ABSTRACT

This dataset is a compilation of Chinese transcriptions of Buddhist terms produced by translators from the late Hàn period. It is a compilation of the previous works of Coblin (1983), Karashima (2010), Vetter (2012), Hill, Nattier, Granger, and Kollmeier (2020) for the Chinese transcriptions. To these were added phonological reconstructions of the Chinese terms for late Hàn from Schuessler (2009) and Middle Chinese from Baxter and Sagart (2014), as well as the Gandhari equivalents of Sanskrit and Pāli terms from Baums and Glass (2002). This dataset, shared on Zenodo, aims at being the new state-of-the-art dataset on Buddhist transcription material and can be used by anyone working on Hàn Chinese phonology and will help better understanding the possible language sources of the Chinese transcriptions, as well as the phonology of the target Chinese dialects.

CORRESPONDING AUTHOR:
Julien Baley
Department of East Asian Languages and Cultures, SOAS University of London, London, UK
julien.baley@gmail.com

KEYWORDS:
Chinese phonology; Buddhism; Hàn dynasty; Ân Shigāo; Lokakṣema; Kâng Mèngxiáng

TO CITE THIS ARTICLE:
(1) OVERVIEW

REPOSITORY LOCATION
https://doi.org/10.5281/zenodo.8115154

CONTEXT

While Modern Chinese is known for its short words, its simple syllable structure, and its tones, in the distant past Chinese was a very different language; Old Chinese (1300–100 BCE) lacked tones, had complex syllable structure with consonant clusters, and used prefixes and suffixes to form new words. By the early 7th century, when the earliest extant Chinese pronunciation dictionary was published, Middle Chinese was already recognizably a form of the language we know today.

It was during the 漢 dynasty (206 BCE–220 CE) that the radical transition between those two stages occurred; it was the first enduring empire in Chinese history, and among the most formative periods for Chinese thought and literature. At this time, the Confucian cultural milieu accompanying classical scholarship thrived. The Confucian classics themselves were edited and (literally) set in stone, while poetry and belletristic prose flourished. The 漢 also saw unprecedented exposure to and influences from foreign cultures, from grapes to backgammon, with Buddhism standing out as the period’s most abiding foreign influence.

As part of the spread of Buddhism from the west, works of Buddhist literature were brought to China and translated by teams of editors (Zücher 2007); amongst them, three figures from the later days of the 漢 dynasty are worth mentioning:

• 安世高 (fl. 148–170), a Central Asian translator active in the Chinese imperial capital of 洛陽, was the first translator of Buddhist texts into Chinese whose name we know (Zacchetti 2019: 630).
• 拉転 (fl. 147–189) was a Buddhist monk from Gandhara also active in 洛陽 (Harrison 2019: 700).
• 康孟詳, of whom little is known, but is generally considered to be born in China from Sogdian parents (Nattier 2008: 102).

These three figures are of particular interest to us here because of their use of transcription in their translations. For instance, while a concept such as धर्म (rule, way, doctrine) ended up being translated into Chinese as *puɑ 法, it can also be found in the translations of Lokäkṣema as *dəm-mɑ 曇摩, a phonetic transcription of a Prakrit word comparable to Pāli dhamma or Gandhari dhaṃma.

The most extensive discussion of the implication of such transcriptions for the phonology of Late 漢 Chinese is Coblin (1983). Since that publication, however, a lot of things have changed: new manuscripts have been discovered and their authorship has been attributed to 安世高 (Zacchetti 2010: 264), providing new transcriptional data, while some other texts traditionally attributed to him have now been classified as later commentaries (Zacchetti 2010: 259–262); our understanding of Old Chinese phonology has dramatically changed and in particular it is now accepted that Old Chinese had a complex syllable structure with consonant clusters in syllable-initial and final position as well as prefixes and suffixes, cf. Baxter (1992), Baxter and Sagart (2014b); finally, our understanding of languages that could have been close to the source languages of the texts being translated by 安世高, Lokäkṣema, and 康孟詳 – in particular Gandhari (Baums 2009) – has progressed.

These developments make it necessary to revisit Coblin’s conclusions regarding the contributions of the Buddhist transcriptional data to our understanding of 漢 Chinese, and the dataset presented here is an attempt to lay out all of the available Buddhist transcriptional data from the Late 漢 period and annotate it with state-of-the-art linguistic knowledge: Sanskrit, Pāli and Gandhari equivalents serve as points of comparison for what the pronunciation of the words might have been in the unknown source language, and Late 漢 Chinese and Middle Chinese reconstructions as illustrations of the transcriptions’ target language.
(2) METHOD

BASE CORPUS

The basis of the dataset is Coblin (1983), whose Buddhist transcriptions include the following texts from the Taishō Tripitaka:

- Ān Shigao
  - T13: Chǎng āhán shì bāōfǎ jǐng 長阿含十報法經
  - T14: Rèn bèn yú shēng jīng 人本欲生經
  - T31: Yì qiě liù shé shōu yín jīng 一切流攝守因經
  - T32: Sì dí jīng 四諦經
  - T98: Pūfā yì jīng 普法義經
- T150A: Zà jīng sīshì piān 雜經四十四篇
- T150A (1): Qǐ chū sān guān jīng 七處三觀經
- T150A (30): Jī gū jīng [經] 梗骨[經]
- T150A (31): Jiǔ hēng jīng [經] 九橫[經]
- T602: Dà ānbān shùyǔ jīng 大安般守意經
- T607: Dào dí jīng 道地經
- Lokāśēma
  - T224: Dàoxíng bārě jīng 道行般若經
  - T280: Dōushā jīng 兜沙經
  - T313: Āchū fuguō jīng 阿闍佛國經
  - T418: Bānzhòu sānméi jīng 殿舟三昧經
  - T458: Wēnshūshì lì wèn púsà shū jīng 文殊師利問菩薩所經
  - T626: Āshēshì wāng jīng 阿闍世王經
- Kōng Mēngxiāng
  - T184: Xiūxíng běnqī jīng 修行本起經
  - T196: Zhōng běnqī jīng 中本起經

ADDITIONS AND REMOVALS

Over the years, scholars have expressed doubts regarding the inclusion of this or that text to the corpus of these translators, while other texts were proposed for inclusion. For the Ān Shigāo corpus, a consensus gradually emerged and is described in detail in Zacchetti (2019), itself based on the work of Zürcher (1977) and Zürcher (1992). Some of the texts in Zacchetti’s list were long considered to be part of Ān Shigāo’s works but were not studied by Coblin. As a result, we added the following texts on top of Coblin’s Ān Shigāo’s corpus:

- T36: Bēnxiāng yǔzhī jīng 本相猗致經
- T48: Shì fā fēi fā jīng 是法非法經
- T57: Lòufēn bù jīng 漏分佈經

---

1 For a detailed discussion of Lokāśēma’s extant corpus, cf. Harrison (1993), and Nattier (2008: 77-85) for a detailed segmentation of the texts into three tiers, each tier representing a level of proximity to Lokāśēma’s own style, and the more distant tiers are posited to be indicative of later revisions of the text.

2 It should be noted that the expected Mandarin reflex for 般若 – as transcribing a word in a Prakrit akin to Pāli pariñā or Gandhari pramāṇa ([pr̥aːnːaː]) – would be bārē; indeed, bārē 般若 had two MC pronunciations, pān and prān, respectively pointing to Eastern Hòn “pān” and “prān”, being good fits for the first syllable of either Pāli pariñā or Gandhari pramāṇa, while re 若 corresponds to MC nyā, pointing to Eastern Hòn “njaʔ”, a good match for ra (ruó 若 points to *njaːk). The standard rendition of 般若 as bārē might suggest a later (hypercorrective) learned reading of 般若 as transcribing the first syllable of Sanskrit prajñā: bō points to Eastern Hòn sp (r) aːj. Nevertheless, as bārē is the defacto standard pronunciation of the word, we use it in the title of T224.

3 While there is a consensus around T13 belonging to Kōng Mēngxiāng’s corpus, one should note that the extant text appears to have undergone later revisions as late as the Eastern Jin dynasty (266–420). See Nattier (2008: 104-105) for a discussion of the external and internal evidence, itself based on Kawano Satoshi (1991).

4 An in-depth discussion of each of the three translators can be found in Nattier (2008): for Ān Shigāo, see Nattier (2008: 41); for Lokāśēma, see Nattier (2008: 75); for Kōng Mēngxiāng, see Nattier (2008: 103).
In addition, T602 Đà nbàn shì yì jìnģ 大安般守意經, originally listed in Coblin (1983) was removed.\(^5\)

For Lokakṣema and Kāng Mèngxiáng, no new texts were added, but for Lokakṣema more transcription words were added from T224 Daoxing bōrê jìnģ 道行般若經, on the basis of Karashima (2010).\(^4\) All the transcription material mentioned so far for the three translators can be found in Hill et al. (2020).

On top of these, two manuscripts discovered in 1999 in the Kongô-ji 金剛寺 temple were ascribed to Ân Shīgăo in Zacchetti (2010: 264); Vetter (2012), in his study of Ân Shīgăo’s lexicon, includes material from the Kongô-ji as well as from T101,\(^7\) and we have retrieved the transcription material from there. The final Ân Shīgăo corpus, starting from Coblin (1983) and applying all the additions and removals, comprises the following texts:  

1. T13: Chăng ânhăn shì bōofâ jìnģ 長阿含十報法經
2. T14: Rên bênh yû shêng jînģ 人本求生經
3. T31: Yì qiê tû shê shòu yìn jînģ 一切流攝守因經
4. T32: Sì dî jîng 四諦経
5. T36: Bênhiăng yûżhi jînģ 本相寂致經
6. T48: Shi fâ féi fâ jînģ 是法非法經
7. T57: Lûu fénbù jînģ 漏分佈經
8. T98: Pûfâ yì jînģ 普法義經
9. T101: Zâ ânhăn jînģ 談阿含経
10. T112: Bâ zhêng dào jînģ 八正道経
11. T150A: Zâ jînģ sîshî piân 難經四十四篇
12. T150A (1): Qiê chûu sàn guân jînģ 七處三觀經
15. T603: Yîn chi rû jînģ 陰持入經
16. T607: Dâo dî jînģ 道地經
17. T1508: Ânhăn kû jî jê shiêr yînyûân jînģ 阿含口解十二因緣經
18. T1557: Áptûn wà fâ xîng jînģ 阿毘曇五法行經
19. Kongô-ji: Ânbân shîuvi jîng 安般守意經 (‘KA’)
20. Kongô-ji: Shiêr mên jînģ 十二門經, Jiê shiêr mên jînģ 解十二門經, and the anonymous commentary (‘TG’)

\(^5\) T605 Chân xîng fâ xiâng jînģ 道行性想經, and T792 Fâ shòu chên jînģ 法受沉經, listed in Zacchetti (2010: 259) as needing to be removed from the Ân Shīgăo corpus, were not listed in Coblin’s work.

\(^6\) Five texts attributed to Lokakṣema in the studies mentioned above are missing from our dataset: T282 (Zhû pûsà qiâ fê bêng yê jînģ 諸菩薩求佛本願經), T283 (Pûsà shî xìng dào pûn 諸菩薩十住行品), T362 (Ámitûsâ sânyésûn fô sîlû jîng toân guoû rênđào jînģ 阿彌陀三藐三佛薩樓樓僅道人經), T624 (Dânxíngbùlôu sôw wên rûlû sâmshî jînģ 通賢無論所問如來法味經), and T807 (Nêi cêng bû ài jînģ 內藏百愛經) since – to the best of our knowledge – no collection of the transcriptions of Indic terms exists. We aim to address this gap in a future publication.

\(^7\) The two manuscripts are nearly identical and contain 4 different texts: Ânbân shîuvi jîng 安般守意經 (‘KA’ in our dataset), Shiêr mên jînģ 十二門經, Jiê shiêr mên jînģ 解十二門經, and finally an anonymous commentary on the two previous texts, collectively ‘TG’ in our datasets.

\(^8\) T101’s status is still a matter of controversy and Hill et al. (2020) chose not to include it. Following Harrison (2002), we have chosen to include it; it contributes 12 new entries to Ân Shīgăo’s corpus, and our dataset is structured to be easy to filter it out if T101 is eventually deemed not to be from Ân Shīgăo.

\(^9\) Vetter (2012) includes T397 (13) Shi jîng pûsà pûn 十方菩薩品 in Ân Shīgăo’s corpus; we follow Nattier (2008: 55–59) who convincingly argues that the text cannot be from the hand of Ân Shīgăo and excludes it.
Altogether, this forms the Chinese basis of our dataset, along with the identification of the corresponding Sanskrit and/or Pāli equivalents. For these, we have relied on the identification made in Vetter (2012) for the Kongô-ji texts and Hill et al. (2020) for the rest.

SOURCE SUMMARY

As a summary, the transcriptions listed in the dataset directly come from the following sources: for Ân Shigâo, we collate Hill et al. (2020), which expands Coblin’s work with more texts and more entries for the existing texts, and Baley (2023), which collects transliteration terms from Vetter (2012) for the Kongô-ji 金剛寺. For Lokakṣema and Kâng Mêngxiâng, we use Hill et al. (2020) (which extends Coblin’s work on Lokakṣema using Karashima (2010)). A comparison of the number of entries between Coblin (1983), Hill et al. (2020), and our dataset, for each translator, can be found in Table 1.

<table>
<thead>
<tr>
<th>TRANSLATOR</th>
<th>COBLIN 1983</th>
<th>HILL ET AL.</th>
<th>NEW DATASET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ân Shigâo</td>
<td>33</td>
<td>33</td>
<td>67</td>
</tr>
<tr>
<td>Lokakṣema</td>
<td>257</td>
<td>280</td>
<td>280</td>
</tr>
<tr>
<td>Kâng Mêngxiâng</td>
<td>54</td>
<td>54</td>
<td>54</td>
</tr>
</tbody>
</table>

INDIC TRANSCRIPTIONS

As the Sanskrit/Pāli information in Hill et al. (2020) was incomplete – for some entries only one of the two languages was provided – we have aimed to complete it where possible; in addition, we have used Baums and Glass (2002) to provide Gandhari equivalents to the Sanskrit/Pāli whenever we were able to identify such equivalents. This will help explore the question of the translations’ source language(s) from a quantitative as well as qualitative point of view. We think that expanding this process to other languages of Central Asia, as their scholarship improves, would be desirable; in particular, we aim to explore Tocharian equivalents in a later project.

CHINESE RECONSTRUCTIONS

We have added columns to provide reconstructions of various stages of Chinese phonology:

- Late Hàn: Schuessler (2007) and Schuessler (2009)
- Middle Chinese: we use the Middle Chinese transcription system (based on the rime books and rime tables) described in Baxter (1992)

(3) DATASET DESCRIPTION

OBJECT NAME

Chinese Transcription of Buddhist Terms in the Late Hàn Dynasty.

FORMAT NAMES AND VERSIONS

OpenDocument Spreadsheet

---

10 In the table, words occurring in multiple places in the corpus of a translator are counted as a single entry (with multiple locations).

11 Of all the transcription words in the 7 extra works added in Hill et al. (2020) compared to Coblin (1983), all were already present in other Ân Shigâo texts, except for Āpâtán 阿毘曇 in T1557’s title which should not be treated as coming from Ân Shigâo because – as an anonymous reviewer suggested – the title is a later addition, and the word itself cannot be found within the text.

12 For Gandhari, we have simply looked up Baums and Glass (2002); for missing Sanskrit / Pāli, we have relied on other entries in the database that had contained the same parts of words; for instance, while Baums and Glass (2002) does not contain an entry for Sanskrit indradatta, it does contain one for indra and datta, and so we have marked the Gandhari equivalent as iṃdra+data, to indicate it is the result of two look-ups.

13 Cf. Boucher (1998) discussing why it might not be possible to prove that the source language is Gandhari.
CREATION DATES
2023-04-01 to 2023-05-06

DATASET CREATORS
Julien Baley, SOAS University of London: Data curation, Investigation, Methodology, Validation.

LANGUAGE
English, Chinese (Late Han, Middle, Modern), Sanskrit, Pāli, Gandhari

LICENSE
Creative Commons Attribution 4.0 International

REPOSITORY NAME
Zenodo

PUBLICATION DATE
2023-06-27

CONTRIBUTING
If you find errors in the dataset, please email the corresponding author.

(4) RE-USE POTENTIAL

By bringing together the scholarly work of many different scholars, this dataset can serve as the basis for further analysis of transcription practices of the Chinese Buddhist translators of the late Han dynasty. For instance, the question of the attributions of translation works is a recurring one and in the case of translators such as Ân Shigao and Lokakṣema – as we have seen – the debate about the authorship of individual texts can take place over many centuries. Our dataset provides a quick reference that can help argue – on internal grounds – whether the transcriptional vocabulary used in a text is typical of a certain translation team and can therefore contribute to discussions of text attributions, including discussions of layering of the translation process.

Another potential re-use of our dataset is to help with interpreting Gandhari texts: a good number of the texts included in the present dataset are translations of texts that are no longer extant; with new excavations of manuscripts in Gandhari and other languages, as well as the gradual cataloguing of the existing ones, our dataset of equivalence between Chinese and Gandhari may help – in the future – to identify the source text of such translations or – since the editorial history of such texts is generally more complicated – at least to identify passages that bear similarities to our known Chinese texts and help interpret the Gandhari manuscripts and our understanding of the doctrinal development underlying the diffusion of such texts.

Finally, as the dataset contains Chinese transcriptions of Buddhist concepts and their equivalents in several languages, this information can be used to try and qualify the source language of those transcriptions. For example, does a given Chinese transcription of a Buddhist term show greater similarity to its equivalent in Sanskrit, Pāli, Gandhari or yet another language, and what does it tell us about the likely phonetic characteristics of the translation’s source language?

In the earlier example of dharma transcribed by Lokakṣema as *daṃ-ма, the reconstruction of a final *-m is certain for *daṃ, this seems to exclude the possibility of a transcription from Sanskrit dharma, and instead the choice of two syllables, the first ending in *-m and the second starting with *m- would indicate a gemination in the source language, as is for instance found in Prakrits such as Pāli damma and Gandhari dharmma.

Following such analysis at the corpus level, does a trend emerge from all the transcriptions from a certain translator or translator team? For instance, one may notice in Ân Shigao’s transcriptions a certain trend for sibilants to match Gandhari better than Sanskrit or Pāli, as illustrated in Table 2, while Lokakṣema – who was from Gandhara – shows more variation in
his transcriptions: some words match more closely Pāli models, as in his use of *ʔa ṭṣan dai 阿旃陀 that better matches Pāli accanta than Skt. atyanta or Gdh. acada,14 while others show a Gandhari slant, such as *ṭṣan diei羼提 being closer to Gandhari ks̄̄m̄ti15 than to Pāli khanti.16

Conversely, the parallel question can also be investigated: given the Chinese transcriptions, what can one learn about the dialect of Chinese spoken by the translator team? What phonological features of that dialect can be discovered from the choice of Chinese characters to transcribe certain syllables of the original Buddhist term? Such questions are of extreme importance to the reconstruction of the historical development of Chinese phonology during the late Hàn period.

ACKNOWLEDGEMENTS

We would like to extend our thanks to Jan Nattier for sharing her data and the advice she has provided as part of this project, and an anonymous reviewer of this article for pointing out important gaps in our initial submission.

FUNDING STATEMENT

This work was funded by the Arts and Humanities Research Council (AHRC), UKRI, as part of the project “Han Phonology: When Chinese Became Chinese.” Project Reference: AH/V008722/1. Principal Investigator: Ernest Caldwell, SOAS University of London. Co-Investigator: Nathan Hill, SOAS University of London, Trinity College Dublin.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

Julien Baley: Data curation, Formal Analysis Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. Nathan Hill: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing. Ernest Caldwell: Funding acquisition, Project administration, Supervision, Writing – review & editing.

AUTHOR AFFILIATIONS

Julien Baley orcid.org/0000-0003-1056-6211
Department of East Asian Languages and Cultures, SOAS University of London, London, UK

Nathan Hill orcid.org/0000-0001-6423-017X
Department of East Asian Languages and Cultures, SOAS University of London, London, UK; Trinity Centre for Asian Studies, Trinity College Dublin, Dublin, Ireland

Ernest Caldwell orcid.org/0000-0003-3390-3751
Department of East Asian Languages and Cultures, SOAS University of London, London, UK


15 Gandhari <kṣ̄>⇒[tʂ]. (Baums, 2019: 7)

16 Again, the actual source language of the transcriptions might have been yet another language, and the examples are provided for illustration only; further investigation is required.
REFERENCES


TO CITE THIS ARTICLE:


Submitted: 08 May 2023
Accepted: 04 July 2023
Published: 21 July 2023

COPYRIGHT:

© 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

Journal of Open Humanities Data is a peer-reviewed open access journal published by Ubiquity Press.