Check for updates

DATA NOTE

A database of qù-tone alternations in Ancient Chinese

[version 1; peer review: 3 approved with reservations]

Barbara Meisterernst[1], Johann-Mattis List[2]

[1]Institut für Linguistik, Universitat Stuttgart Institut fur Literaturwissenschaft, Stuttgart, Baden-Württemberg, 70174, Germany
[2]Chair for Multilingual Computational Linguistics, University of Passau Faculty of Arts and Humanities, Passau, Bavaria, 94032, Germany

v1 **First published:** 22 Oct 2025, **5**:326
https://doi.org/10.12688/openreseurope.21142.1
Latest published: 22 Oct 2025, **5**:326
https://doi.org/10.12688/openreseurope.21142.1

Abstract

Alternations in the entering tone (*qù*-tone) in Ancient Chinese have for a long time fascinated scholars, since they seem to give hint to relics of morphology in the history of Chinese, contrasting strongly with the isolating structure of all modern varieties of Chinese. Here we present a transparently assembled collection of entering tone alternations in the history of Chinese, derived from Lù Démíngs historical annotation of the classics, the Jīngdiǎn Shìwén, which gives early hints on such alternations by means of historical *fǎnqiè* spellings. The database is available in two flavors. On the one hand, it can be accessed through a web interface that allows interested users to browse through the data in linked form. On the other hand, the data is available in the formats recommended by the Cross-Linguistic Data Formats initiative. This format does not only offer quick programmatic access to experienced users, but is also specifically apt for the purpose of archiving the data. In the study, we illustrate how the data was assembled and curated, and illustrate how the data can be put to concrete use.

Plain language summary

It is commonly believed that words in Old Chinese that show alternations only in their tones, might give hints to former grammatical processes in the history of the Chinese language. The study presents a database in which such alternations have been assembled for alternations involving the so-called entering tone, one of four tones in Old Chinese. The database allows to browse the evidence in various forms. The study presents how the data were assembled and how the online interface to the database was created.

Keywords

Chinese linguistics, Chinese traditional phonology, qù-tone in Ancient Chinese, Chinese tones, database, Cross-Linguistic Data Formats

Open Peer Review

Approval Status ? ? ?

	1	2	3
version 1 22 Oct 2025	? view	? view	? view

1. **Christoph Harbsmeier**, University of Oslo CEES, Oslo, Norway
2. **Zev Handel**, University of Washington, Washington, USA
3. **Daniel Kadar**, Dalian University of Foreign Languages, Dalian, China

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the Horizon 2020 gateway.

This article is included in the Languages and Literature gateway.

This article is included in the Horizon Europe gateway.

This article is included in the European Research Council (ERC) gateway.

This article is included in the Linguistic Diversity collection.

Corresponding authors: Barbara Meisterernst (barbara.meisterernst@gmail.com), Johann-Mattis List (mattis.list@uni-passau.de)

Author roles: Meisterernst B: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **List JM**: Software, Validation, Visualization, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: This research was funded by the DFG project 143118582 (BM) and the the ERC Consolidator Grant 101044282 (JML). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency (nor any other funding agencies involved). Neither the European Union nor the granting authority can be held responsible for them.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Meisterernst B and List JM. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Meisterernst B and List JM. **A database of qù-tone alternations in Ancient Chinese [version 1; peer review: 3 approved with reservations]** Open Research Europe 2025, **5**:326 https://doi.org/10.12688/openreseurope.21142.1

First published: 22 Oct 2025, **5**:326 https://doi.org/10.12688/openreseurope.21142.1

Introduction

Minimal pairs involving tonal alternations ('Derivation by tone change') of roots in one of the three tones of Middle Chinese (even tone *pīngshēng* 平聲 A(1), rising tone *shǎngshēng* 上聲 B(2), entering tone *rùshēng* 入聲 D(4)) with the *qùshēng* 去聲 (Tone C(3)) in Middle Chinese are quite common in Archaic Chinese. Haudricourt (1954) was the first who proposed that the emergence of the *qùshēng* was caused by the loss of a suffix *-s (Haudricourt, 1954; Sagart, 1999), which changed into *-h and further into the *qùshēng*. The process probably took place in different stages during the Late Archaic (5th – 3rd c. BCE) and Early Middle Chinese periods (2nd c. BCE – 3rd c. CE). The reconstructed suffix *-s is the most generally accepted derivational affix of Archaic Chinese and it has been connected to different morphological processes in the subsequent literature. Numerous functions have been attributed to the suffix (e.g. Downer, 1959; Haudricourt, 1954; Jin, 2006; Sagart, 1999; Sun, 2015; Unger, ms; Wang, 2014; Wuyun, 2024) with eight major categories proposed in Downer (1959, see also Jacques, 2016), which have been reduced subsequently to two major functions, (a) a valency decreasing function, connected to passive (Schuessler, 2007; Xing & Schuessler, 2020), perfective (Jin, 2006), and resultative (telicity) values (Meisterernst, 2023), and (b) a valency increasing (causative) function (e.g. Schuessler, 2007; Wang, 2014; Wuyun, 2024; Xing & Schuessler, 2020). The investigation of these functions in verbal derivations are at issue in the present project.

The oldest and one of the most important sources for a reliable investigation in the tones of Archaic Chinese is the *Jīngdiǎn Shìwén* 經典釋文 (JDSW) by Lù Déming 陸德明 (550–630 CE), completed before 589, which consists of a collection of semantic and phonological glosses from the earlier commentary literature (Branner, 2003). Different from the traditional commentaries on the Classical Chinese literature in which the running text is interspersed with different kinds of glosses and commentaries, Lù Déming did not attempt to subordinate his thousands of glosses to a running text but instead only commented on singular words and phrases he cited from the Classical Literature. The texts he commented on overlap with those collected in the Thirteen Confucian Classics (*Shísān jīng* 十三經), leaving out the text by Mencius, but including two Daoist Classics, the *Dàodé jīng* 道德經 and the *Zhuāngzǐ* 莊子. Whereas earlier commentaries and sources, such as the character dictionary *Shuōwén jiězì* 說文解字 (1st to 2nd c. CE), mostly used homophonic glosses or the *dúruò* glosses (using the schema X 讀若《丙》) "read [this character] as X"), Lù Déming used the so-called *fǎnqiè* 反切 glossing system, which developed from at least the 3rd c. CE on, based on Hàn period and later glosses to the Classics. In the *fǎnqiè* system a character is divided into an initial and a final (rime) part (including tones) which are each represented by a single character, as illustrated in Table 1 (see also Branner, 2000 and List, 2018).

Scholars such as Downer (1959), Huang (1992), Wang (2014), Sun (2015), (Wuyun, 2024) based their research on the JDSW for their investigations of different aspects of the morphological functions of the suffix *-s, but frequently studies on the morphology of Chinese use different, including younger sources such as the Sòng period dictionary *Guǎngyùn* 廣韻, for their research. Additionally, the methods employed in research are not always transparent and the situation in Archaic Chinese is such that the establishment of cognate sets is faced with numerous challenges, since problems ranging 'from allographs (*yìtǐzì* 異體字), phonetic loans (*tōngjiǎzì* 通假字), to lexical borrowings' have to be considered (Pulini & List, 2024a). One example illustrating this problem involves different graphic representations of the lexical root shì 視, OC *gijʔ 'to see' and its derivation, shì 示 OC *s-gijʔ-s 'to show' (Baxter & Sagart, 2014: 56, cf. Pulini & List, 2024a: 2024).

In the present project all *qùshēng* roots involving a verbal reading, and all roots which alternate with a (verbal) cognate (a derivation) in the *qùshēng* are at issue. We include also those cases where a deverbalized noun appears in the *qùshēng*. The objective is to contribute to the research on the basic function(s) of the suffix *-s based on the Archaic Chinese instances glossed in the JDSW and the impact their loss had on the diachronic development of the Chinese syntax from a more synthetic to a more analytic system (Huang & Roberts, 2017). The JDSW data is the starting point of our investigation, since it provides the earliest data reliably representing differences in pronunciations, in order to clarify the existence of a clear morphological and derivational relation between cognates (Fellner & Hill, 2019) as far as possible. In relying on this very early source, we try to consolidate the considerable disagreement in the reconstruction of affixes in Chinese. For instance, our data show that the reconstructions of the Archaic Chinese affixes in the phonological literature do not always correspond to the pronunciation glosses in the JDSW (see also Jacques, 2022). This may be due to the use of different sources by the respective scholars for their reconstructions (for an overview on sources used in the reconstruction of Archaic Chinese see Baxter and Sagart, 2014: 2f). The collection of data so far also shows considerable differences in the *fǎnqiè* annotations of the JDSW and of the *Guǎngyùn* respectively, suggesting that the system of sound glosses underwent significant changes towards a certain degree of standardization from the Sui to the Song period. The differences between the JDSW and the Guǎngyùn systems provide some evidence for a genuine early system of sound glosses representing morphological differentiations contrary to the assumptions of the Chinese Purist School of the *Kǎozhèng* 考證 philologists of the Manchu period. The latter were opposed to the tradition

Table 1. The four different tones represented by fǎnqiè spellings.

堂 táng[th]	EMC	daŋ:	徒郎	d [d] [ɔ l]	aŋ (MM: t(ú l)áng)
蕩 dàng[t]	EMC	daŋʻ:	徒朗	d [d] [ɔ l]	aŋʻ(MM: t(ú l(ǎng)
宕 dàng[t]	EMC	daŋh:	徒浪	d [d] [ɔ l]	aŋh (MM: t(ú l)àng)
鐸 duó [t]	EMC	dak:	待各	d [d] [əjʻk]	ak ʻ(MM: d(ài g)è)

of the variant readings collected in 'rime-books' and standard commentaries from the Han period on. The ideal of the purist school, whose scholars only accepted very few of the proposed tonal alternations, was the conception of the Chinese writing system in which one characterstood for only one reading (Branner, 2003).

In a first step, the data collection and the resulting database is intended to enhance the transparency of work with the sound glosses in the JDSW, which are presented with their Middle Chinese pronunciations. Similar to the annotation of ancient Chinese documents written on Bamboo Slips (Pulini & List, 2025) a list of *fǎnqiè* with Middle Chinese pronunciations linked to the character and the respective occurrences in the JDSW via an ID could serve as a basis for a more consistent system of reference in the research on the historical morphology of Chinese.

Materials and methods

Materials

Initial data collection was done in tabular form, created as an Excel spreadsheet. The starting point for the materials collected manually so far was the digital version of Unger's (1989) *Glossar des Klassischen Chinesisch*, a wordlist with 3500 entries. From this a list of 956 words with a *qùshēng* reading have been extracted into an excel file, i.e. ca. 27% of the entire lexicon assembled in the *Glossar*. According to a first survey, 562 of these 956 roots (58,79%) involve a verbal reading. This list of verbal roots has been checked against Baxter & Sagart (2014), Pulleyblank's Lexicon of Reconstructed Pronunciations (1991), the monographs of Jin (2006) and Sun (2007), and an article by Wuyun (2024). All words which have a verbal meaning as one of their lexical entries in the respective sources were added to the spreadsheet. In a later step the reconstructions of Old Chinese by Schuessler (2007) and the *fǎnqiè* from the *Guǎngyùn* with their MC reconstructions (taken from Pulini & List, 2025) were included. From the two monographs and the article by Wuyun (2024) examples were additionally incorporated in the spreadsheet. The resulting sheet 'Verbs_complete' contains 1468 entries and has the structure indicated in Table 2.

All entries marked for tonal alternation were extracted to a separate table, which subsequently was checked against the JDSW. In a first step we specifically searched for those entries which appear with one or more *fǎnqiè* involving tonal alternation in the JDSW, leaving out the original (non-derived) readings marked by *rú zì* 如字 ('like the character') and the *yīn* 音 X ('sounds like X') glosses. All *fǎnqiè* were listed, including those that were homophonous or near homophonous. The total number of tone alternation roots from the 'Verbs_complete' sheet checked against JDSW is 547, of those 46 (8,41%) do not appear with a FQ in JDSW and 33 (6,03%) are not attested at all. Of the remaining roots, 496 were included in the *Qù-Tone Alternations* database.

Apart from the manually compiled spreadsheet, digital versions of additional databases were used as provided by their publishers or as they could be retrieved by us. These include *fǎnqiè* spellings and automatically generated Middle Chinese readings from the *Guǎngyùn* resource provided by Pulini and List (2024b), Baxter and Sagart's Old Chinese readings in digital form as provided by the authors, a digital version of Schuessler's (2007) Old Chinese reconstruction as provided in digital form by List *et al.* (2017), and a digital version of the *Jīngdiǎn shìwén* (Lu 583[1985]).

Methods

Creation of the basic spreadsheet. The data for the 'initial sheet' file, the basis for the first version of the Qù alternations database containing *fǎnqiè* from the JDSW was assembled and commented on manually based on the digitized version of Unger's (1989) *Glossar*. In a second step Pulleyblank (1991) and the appendix of Baxter and Sagart (2014) were checked for additional lexical items which display possible connections with a verbal root. Those roots were entered manually into the extended spreadsheet together with the respective English glosses provided in Pulleyblank (1991) and Baxter and Sagart (2014), and with the German glosses in Unger's (1989) *Glossar*. On the basis of the dictionary entries a preliminary classification of the parts of speech was added to the table for easier analysis of the data. This allows the automatic exclusion of proper names and of nouns, and other lexical items which do not display any possible connections to verbal roots. The relevant data from Jin (2006), Sun (2007), and Wuyun (2024) was added manually with the help of a research assistant. In a third step, the Old Chinese readings of Schuessler (2007), and the *fǎnqiè* from the *Guǎngyùn* with their Middle Chinese readings were added automatically, using the method by Pulini and List (2024a).

From the initial sheet file all entries marked for tone alternations were extracted automatically into a separate spreadsheet on tonal alternations in the JDSW which includes the following columns:

Character; Pīnyīn; possible alternations of initials, vowels, medials; fǎnqiè of tones A, B, C, D respectively, followed each by the Middle Chinese reconstruction generated by the Pulini and List's method. If the Middle Chinese reading was not available or incomplete, an attempt to reconstruct the reading following Baxter and Sagart (2014) was undertaken manually; these readings are stored in a separate table. The JDSW qù alternation table also contains comments in the JDSW regarding double readings, and occasionally comments on scribal errors and other problems from later sources. For each tone, lists of the examples from the Archaic literature commented on in the JDSW were collected and checked against the Academia Sinica Corpus (https://ancientchinese.ling.sinica.edu.tw/ASACC_index/), the CTEXT Corpus (https://ctext.org), and the Scripta Sinica database of Academia Sinica (https://hanchi.ihp.sinica.edu.tw/ihp/hanji.htm).

Integration of data in cross-linguistic data formats. In order to make for a more detailed integration of the original spreadsheet with additional digital resources, we decided to model the data in CLDF. CLDF is a standard for tabular data, recommended by the Cross-Linguistic Data Formats initiative (Forkel *et al.*, 2018). While CLDF originally targets cross-linguistic

Table 2. Structure of sheet "Verbs Complete" in the initial spreadsheet.

Column	Name	Comment
A	Character	Chinese character, sometimes including variant forms
B	Reconstruction Schuessler	Schuessler's (2007) Old Chinese reconstructions
C	Middle Chinese	Created by edictor or manually
D	Guǎngyùn fanqie	Fǎnqiè as taken from the Guǎngyùn
E	JDSW fanqie	Created by edictor or manually
F	Pīnyīn	Mandarin Chinese transliteration
G	Reconstruction MC Unger	Middle Chinese reconstruction by Unger (1989)
H	Reconstruction EMC Pulleyblank	Pulleyblank (1991)
I	Baxter & Sagart MC	Middle Chinese, Baxter & Sagart, 2014
J	Baxter & Sagart OC	Old Chinese, Baxter & Sagart, 2014
K	Reconstruction Jin (2006)	Reconstruction of Old Chinese
L	English gloss	Pulleyblank (1991); translation of Unger (1989)
M	German gloss	Unger (1989)
N	Tone alternation yes/no	Indication of the presence of tone alternations
O	Baxter & Sagart gloss	Baxter & Sagart (2014)
P / Q	alternative glosses of non-tone C variant	Baxter & Sagart (2014); Pulleyblank (1991); Unger (1989)
R	Tentative syntactic definition	Definition as state Verb, telic Verb, resultative Verb, Noun, etc. based on lexicon entries
S	Ref. Jin (2006)	Ref. to the discussed lexical item
T	Example Jin (2006)	Representative example for the discussed lexical item
U / V / W	comments	On problematic cases etc.
X	Reconstruction EMC Pulleyblank	Reconstructions of Early Middle Chinese of the non-tone C word
Y	Sun (2007) lexical root	Based on the distinction in Sun (2007) into the original (root) and derived form of a word
Z	Sun (2007): derived form	Based on the distinction in Sun (2007) into the original and derived form of a word
AA / AB / AC / AD	Wuyun (2024) examples and comments	Examples from the Wuyun (2024) article on the proposed causative functions of the suffix *-s and brief comment on their discussion

datasets, the basic format can also be adapted for monolingual datasets and has been used to model Old Chinese texts written on bamboo slips with additional annotations (Pulini & List, 2025). The advantage of using CLDF for data representation are two-fold. On the one hand, CLDF allows for a facilitated integration of individual datasets from different sources, thanks to CLDFBench (https://pypi.org/project/cldfbench, Forkel & List, 2020), a Python package that facilitates the creation of CLDF datasets by allowing users to mix custom Python code with standardized predefined functions that work across multiple platforms. On the other hand, datasets, once provided in CLDF, can be easily published online with the help of the CLLD framework (https://clld.org, Forkel, 2014), a framework that facilitates the exploration of tabular data in websites, offering a common look-and-feel specifically to linguists, who know the framework well due to the success of large databases like the World Atlas of Language Structures (https://wals.info, Dryer & Haspelmath, 2013) or the Database of Cross-Linguistic Colexifications (https://clics.clld.org, Rzymski *et al.*, 2020). Furthermore, an extra bonus of creating CLDF datasets, consists in extra tests on data integrity that will be automatically carried out when aggregating data from different sources.

In order to create the integrated Database of Qù-Tone Alternations, we first designed a data model that links the different types of data to each other, and then created a Python script that would allow us to carry out the aggregation automatically with the help of the CLDFBench package. The data model consists of two tables, one entry table and one table for text examples. The entry table provides information on all the characters that were considered in the original spreadsheet, with additional information on *fǎnqiè* spellings, Middle Chinese and Old Chinese readings, and further information. The example table lists all automatically inferred occurrences of the characters in question in the JDSW, with automatically inferred *fǎnqiè* spellings. Rows in the entry table are linked to one or more IDs in the example table, accounting for the fact that several characters occur with multiple glosses and *fǎnqiè* spellings in the JDSW.

Creation of the CLLD web application. Once a CLDF data set has been created, the deployment of the data in the form of a CLLD web application is straightforward. The CLLD application allows to investigate the data from three perspectives through three separate tables. The Roots table lists individual characters along with all attested readings, mimicking the original structure of the spreadsheet, but with entries linked to detailed views of the data. Roots — representing root morphemes in Chinese that can surface in different forms, forming individual words — are linked to their Occurrences, which are in turn disambiguated through different Middle Chinese readings derived from different *fǎnqiè* spellers. Occurrences themselves link to the JDSW Reference. Here, individual occurrences of identical *fǎnqiè* spellers can in turn recur several times in the JDSW. The application allows for a convenient browsing of the three tables, given that they are interlinked and can additionally be filtered by all their individual fields.

Dataset validation

Detailed data inspection

In order to illustrate the benefits of the data curation and data integration steps that we have followed to create the Database of Qù-Tone Alternations, it is useful to inspect the database itself. Figure 1 shows the Roots table, the first table offered by the web application (https://qualternations.digling.org/parameters). Here, each character is linked to one or more (1) readings in Pīnyīn transliteration, (2) tonal values (using numbers to denote the four tones in Middle Chinese and Old Chinese), (3) alternation types (listing Chinese names in Pīnyīn for the four tones), (4) Middle Chinese readings (computed using the approach by Pulini & List, 2024b), and (5) the number of occurrences with different *fǎnqiè* spellers. The Roots table is filtered by the column Readings, listing only entries that start with *bj*.

When clicking on any of the identifiers in the column ID of the roots table, this will provide a filtered view of the entire Occurrences table of the database, listing only the occurrences of the character with the corresponding ID, as shown in Figure 2. Here, different *fǎnqiè* spellers with potentially resulting different readings in Middle Chinese can be contrasted with each other for individual characters.

Clicking on the ID in the ID column a further time will show the occurrences of a character with its particular *fǎnqiè* in the JDSW. This table view, illustrated in Figure 3, allows to check the original entry or the original entries and comments provided by the resource.

General statistics on the data

The web interface with its filtering facilities also allows for a convenient — manual — inspection of major statistics on *qù*-tone alternations assembled from the JDSW. While this could also be directly computed from the CLDF representation of the data, with the help of Python scripts, the filtering functionalities allow for a more flexible, direct interaction with the data than a script run on the terminal of a computer, given that the results can be immediately inspected.

In Table 3, we have listed the absolute frequencies along with their proportion with respect to the total number of roots for



Figure 1. Screenshot of the Roots table in the database.



Figure 2. Occurrences table of the web application, listing entries corresponding to the character fàn 汎 with ID 196.

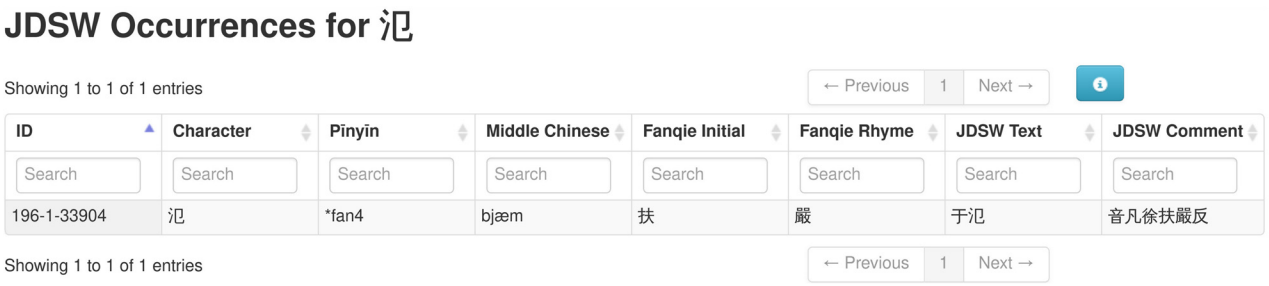

Figure 3. Occurrence of the character fàn 汎 with fǎnqiè 孚劒 in the JDSW.

Table 3. Alternation types by frequency, inferred from the data presented in the web application. Alternations constituting more than 10% of the data are highlighted.

Alternation Type	Frequency	Proportion
1	31	6.25
1 2	9	0.94
1 3	116	12.18
1 4	1	0.12
1 2 3	25	2.99
1 2 4	1	0.12
1 3 4	5	0.62
1 2 3 4	4	0.5
2	23	2.88
2 3	79	10.17
2 4	1	0.14
2 3 4	3	0.43
3	122	17.58
3 4	60	10.49
4	16	3.13
Total	496	100

all observed alternation types. *Qù*-tone alternations constitute the majority of observed alternations here. While this is not surprising — given that the objective of the Database of Qù-Tone Alternations are alternations involving the falling tone — we can see that the proportions of alternations with respect to other tones does not show any particular trend in the data. This coincides well with the hypothesis that alternations involving the falling tone were derivational in nature, and would thus apply to any root, independent of its original phonology. What can also be seen from the general statistics is that — with a total of 292 cases involving *qù* tone alternations — there is substantially more evidence for this alternation type than was previously discussed in the literature. This warrants a closer investigation in future work.

Conclusion and outlook

With this brief study we have presented the digital version of a number of *fǎnqiè* from the earliest collection of pronunciation glosses, the JDSW, as the basis for a project on the diachronic development of the Chinese language. The specific objective of this database is to facilitate the research on the morphological functions of the reconstructed suffix *-s of Archaic Chinese, which developed into the *qùshēng* in Middle Chinese. Tonal alternations in Middle Chinese resulting from earlier affixation have been comprehensively discussed in the literature since Haudricourt (1954). The suffix *-s is the most generally accepted affix of Archaic Chinese and its morphological and syntactic analysis together with cross-linguistic studies on the derivational morphology in the Tibeto-Burman

languages have gained considerable interest in recent years (e.g. Jacques, 2016; Zhang, 2022). The Qù alternations database presented in our study so far facilitates the analysis of the statistical distribution of the tonal alternations collected in the JDSW in providing the attested FQs for each character together with their Middle Chinese pronunciations.

In a next step, the reconstructions of Old Chinese by Schuessler (2007) and Baxter and Sagart (2014) and the English glosses from the literature (Baxter & Sagart, 2014; Pulleyblank, 1991) collected in the Verbs_complete excel table will be added to the database. Additionally, the database will be expanded to not only including those words which display tonal alternations, but all *qùshēng* words from the Verbs_complete table which have a verbal as one of their possible readings (around 950 more entries). This will result in a digital dictionary of Archaic Chinese lexical roots in the *qùshēng* (original or derived) with verbal (but not necessarily only verbal) readings with their Middle Chinese pronunciations based on the JDSW and the *Guǎngyùn* and with the original *fǎnqiè* from the JDSW. A number of these verbs display more than one alternation, i.e., alternations of the initial, the medial, and/or the vowel together with the *qùshēng*. A digitized version of the data allows for easy comparison of different annotation systems in the sound glosses and for philological and linguistic research on the historical morpho-phonology of Chinese.

Ultimately, in a last step, results of the research on the morpho-syntactic functions of the *-s suffix (the *qùshēng* in Middle Chinese) and the diachronic morphosyntax of Chinese will be linked to the database. The most representative examples from the Archaic Chinese literature for the proposed functions of the *s-suffix, e.g., as a marker of telicity or resultativity, causativity, nominalization etc., will be linked to the respective verbs. The examples will be presented with transcriptions in Pīnyīn, with interlinear glosses following the Leipzig glossing rules, and with translations. The glosses for the verbs discussed will include the Middle Chinese pronunciations based on the earliest FQ from the JDSW and the reconstructions of their Old Chinese forms following Baxter and Sagart (2014). The Digital Edition of the Cáo Mò Zhī Zhèn 曹沫之陳 (https://cmzz.digling.org/) will serve as a model for the linking and the representation of the examples. In this way, the final database will allow to query different aspects of Archaic Chinese such as the lexicon, the phonology, morphology, and syntax with corpus linguistic tools. Chinese characters and their *fǎnqiè* are transcribed in Pīnyīn, with their Middle Chinese pronunciations and Old Chinese reconstructions, and glossed in English; the examples for the syntactic structure of the verbs discussed will include English and linguistic glosses and translations. This way the digitally computational data will be made accessible also for linguists not familiar with (Archaic) Chinese, who are interested in the research on cross-linguistic diachronic morpho-syntax. An open access digitized version of the data and the ensuing research results as has been proposed here will enhance the transparency of the research on the historical morphology of Chinese and make it available to a greater linguistic community.

Ethics and consent

Ethical approval and consent were not required.

Data availability

Data available from: https://codeberg.org/calc/qualternations

Archived data available from: https://doi.org/10.5281/zenodo.17244146 (Meisterernst & List, 2025)

License: CC-BY-4.0

Software availability

Source code available from: https://codeberg.org/calc/qualternations

Archived code available from: https://doi.org/10.5281/zenodo.17244146 (Meisterernst & List, 2025)

License: CC-BY-4.0

Reporting guidelines

Data and code are shared with a permissive CC-BY-4.0 License.

Acknowledgements

We thank Robert Forkel for providing tips with CLDF data handling and CLLD deployment.

Supplementary material

Data and code are curated on Codeberg (https://codeberg.org/calc/qualternations, Version 0.1) and archived on Zenodo (https://doi.org/10.5281/zenodo.17244146). Data and software are both provided in this repository. The web-application can be accessed from https://qualternations.digling.org, the code that underlies this web-application is also curated on Codeberg (https://codeberg.org/digling/qualternations-clld, Version 0.1).

References

Baxter WH, Sagart L: **Old Chinese. A new reconstruction**. Oxford University Press: Oxford, 2014.
Reference Source

Branner DP: **The rime-table system of formal Chinese phonology**. In: Sylvain Auroux, E. F. K. Koerner, Hans-Josef Niederehe, and Kees Versteegh (eds): *History of the language sciences.* De Gruyter: Berlin and New York, 2000; 46–55.
Publisher Full Text

Branner DP: **On early Chinese morphology and its intellectual history.** *J R Asiat Soc.* 2003; **13**(1): 45–76.
Publisher Full Text

Downer GB: **Derivation by tone-change in classical Chinese.** *Bulletin of the School of Oriental and African Studies.* 1959; **22**(1): 258–290.
Publisher Full Text

Dryer MS, Haspelmath M: **World atlas of language structures online**. Max Planck Institute for Evolutionary Anthropology: Leipzig, 2013.
Reference Source

Fellner HA, Hill NW: **Word families, allofams, and the comparative method.** *Cahiers de Linguistique Asie Orientale.* 2019; **48**(2): 91–125.
Publisher Full Text

Forkel R: **The cross-linguistic linked data project**. In: *Proceedings of the 3rd Workshop on Linked Data in Linguistics: Multilingual Knowledge Resources and Natural Language Processing.* 2014; 61–66.
Reference Source

Forkel R, List JM, Greenhill S, *et al.*: **Cross-linguistic data formats, advancing data sharing and re-use in comparative linguistics.** *Sci Data.* 2018; **5**: 180205.
Publisher Full Text

Forkel R, List JM: **CLDFbench. Give your cross-linguistic data a lift**. *Proceedings of the Twelfth International Conference on Language Resources and Evaluation*. 2020; 6997–7004.
Reference Source

Haudricourt AG: **Comment reconstruire le chinois archaïque [reconstructing ancient Chinese].** *Word.* 1954; **10**(2–3): 351–364.
Publisher Full Text

Huang CTJ, Roberts I: **Principles and parameters of universal grammar**. In: Roberts, Ian (ed.). *The Oxford Handbook of Universal Grammar*. Oxford University Press: Oxford, 2017.
Publisher Full Text

Huang黃 K坤堯: **Jīngdiǎn shìwén dòngcí yìdú xīntán 經典釋文動詞異讀新談. [New survey on the variant pronunciation of verbs in the Jingdan Shiwen]**. Xuesheng shuju: Taibei, 1992.

Jacques G: **How many *-s-suffixes in Old Chinese?** *Bull Chin Linguist.* 2016; **9**: 205–217.
Publisher Full Text

Jacques G: **On the nature of morphological alternations in Archaic Chinese and their relevance to morphosyntax.** *Bulletin of the School of Asian and Oriental Sciences.* 2022; **85**(3): 475–494.
Publisher Full Text

Jin L: **Shànggǔ Hànyǔ xíngtài yánjiū 上古漢語形態研究[Morphology in Archaic Chinese]**. Hefei, 2006.

List JM: **More on network approaches in historical Chinese phonology (音韵学).** In: The 2nd Li Fang-Kuei Society Young Scholars Symposium. 2018; 157–174.
Reference Source

List JM, Pathmanathan J, Hill N, *et al.*: **Vowel purity and rhyme evidence in old Chinese reconstruction.** *Lingua Sinica.* 2017; **3**(1): 5.
Publisher Full Text

Lu陸 D德明: **Jīngdiǎn Shìwén. 經典釋文[Explanation of the Classics]**. Guji Chubanshe: Shanghai, (583[1985]).
Reference Source

Meisterernst B, List JM: **Qù-tone alternations in the history of Chinese.** Dataset. Version 0.1. Zenodo. 2025.
http://www.doi.org/10.5281/zenodo.17244146

Meisterernst B: **The loss of morphology and the emergence of analytic structures in Chinese.** *J Hist Syntax.* 2023; **7**: 1–55.
Publisher Full Text

Pulleyblank EG: **Lexicon of reconstructed pronunciation in Early Middle Chinese, Late middle Chinese, and Early Mandarin.** UBC Press: Vancouver, 1991.
Reference Source

Pulini M, List JM: **First steps towards the integration of resources on historical glossing traditions in the history of Chinese.** In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*. 2024a; 7343–7348.
Reference Source

Pulini M, List JM: **Finding language-internal cognates in old Chinese.** *Bull Chinese Linguist.* 2024b; **17**(1): 53–72.
Publisher Full Text

Pulini M, List JM: **Using cross-linguistic data formats to enhance the annotation of Ancient Chinese documents written on bamboo slips.** In: *Proceedings of the Second Workshop on Ancient Language Processing*. 2025; 31–37.
Reference Source

Rzymski C, Tresoldi T, Greenhill SJ, *et al.*: **The database of cross-linguistic colexifications, reproducible analysis of cross-linguistic polysemies.** *Sci Data.* 2020; **7**(1): 13.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Sagart L: **The roots of old Chinese.** John Benjamins: Amsterdam, 1999.
Reference Source

Schuessler A: **ABC etymological dictionary of old Chinese.** University of Hawai'i Press: Honolulu, 2007.
Reference Source

Sun孫 Y玉文: **Hànyǔ biàndiào gòucí yánjiū 漢語變調構詞研究 [Investigation of the word-formation by tone-change in Archaic Chinese].** Shangwu Yinshuguan: Shanghai, 2007.

Sun孫 Y玉文: **Hànyǔ biàndiào kǎobiàn 漢語變調考辨 [Examination and distinction of tone-change in Archaic Chinese].** Shangwu Yinshuguan: Shanghai, 2015.

Unger U, Hu好 K古: **Sinologische Rundbriefe.** [Unpublished Manuscript]. (no date).

Unger U: **Glossar des Klassischen Chinesisch [Glossary of Classical Chinese].** Harrassowitz: Wiesbaden, 1989.

Wang王 Y月婷: **Jīngdiǎn shìwén yìdú yīnyì guīlǜ yánjiū "經典釋文"異讀音義規律研究 [Investigation of the patterns of different pronunciations in the Jingdian Shiwen].** Academy of Social Sciences: Beijing, 2014.

Wuyun S: **A Distributed Morphology account of tonal change in Middle Chinese.** *J East Asian Linguist.* 2024; **33**: 489–531.
Publisher Full Text

Xing JZ, Schuessler A: **Semantic extension in old Chinese: direction, transitivity, and voice.** In: Janet Z. Xing (ed.): *A typological approach to grammaticalization and lexicalization.* De Gruyter Mouton: Berlin and New York, 2020; 165–190.

Zhang S: **Rethinking the *-s suffix in old Chinese: with new evidence from Situ Rgyalrong.** *Folia Linguist.* 2022; **43**(1): 129–167.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ? ? ?

Version 1

Reviewer Report 24 January 2026

https://doi.org/10.21956/openreseurope.22870.r66236

© **2026 Kadar D.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Daniel Kadar
Dalian University of Foreign Languages, Dalian, China

I read this manuscript with a lot of interest. Presenting a digital version of a number of *fǎnqiè* from the earliest collection of pronunciation glosses is an important task, and I am happy to recommend the present study for publication. In the following I raise a number of criticisms which are all minor in scope and do not, in my view, warrant another review round.

1. The Authors seem to conflate Introduction with Review of literature. I found it very problematic that the Introduction presents many references, often without appropriately discussing these references. Please resolve this by making the Introduction more focused and moving most of the references to a Review of literature section. Please also avoid chain-quoting, which is a bad practice, such as the following: "(e.g. Downer, 1959; Haudricourt, 1954; Jin, 2006; Sagart, 1999; Sun, 2015; Unger, ms; Wang, 2014; Wuyun, 2024)". If more references are to be mentioned together, such references should be moved into a footnote, but in general chain-quoting is a problematic academic practice.
2. The target audience of this study should include both (!) sinologists and linguists. I suggest to the Authors to revise this manuscript by keeping both these audiences in mind as this would positively interest the visibility and prospective citation figures of the manuscript. In so doing, it would be essential to avoid assuming that readers are familiar with sinological information, such as the sources presented. Rather, it would be useful to tell the Reader more about the particulars of this study, introduce the sources, and so on. Please note that this advice concerns the general study, i.e. not the part where the results are presented in an accessible way.
3. "Materials" is an awkward expression – do the Authors mean Data? Conventionally, Section 3 of a study is called "Data and Methodology" – please consider revising.
4. I really like the Methodology section but one point to improve there is replicability. In linguistics, the general idea behind implementing a certain methodology is that it may not only help the Authors to successfully implement a project but also future scholars to replicate the approach followed by the Authors in the study of other phenomena. This sense of replicability is largely ignored, and consequently the study remains mainly limited to sinology.
5. I suggest that in the Methodology section the Authors insert illustrations as regards how

they database created can be put to use. Later on, illustrations pop up in Section 4 but Section 3 would profit from a general illustration.

In conclusion, I would like to congratulate to the Authors for this excellent project!

Is the rationale for creating the dataset(s) clearly described?
No

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 24 Jan 2026

Johann-Mattis List

These comments are very helpful, many thanks. We will try to revise database and study soon in the nearer future and will by then also provide a point-to-point reply to the comments, which we'll by then hopefully have addressed all.

Competing Interests: No competing interests were disclosed.

Reviewer Report 24 December 2025

https://doi.org/10.21956/openreseurope.22870.r63402

© 2025 Handel Z. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Zev Handel
University of Washington, Washington, USA

****

As the authors correctly note, a searchable, sortable dataset of qù-tone alternations in the medieval Chinese text *Jīngdiǎn Shìwén* 經典釋文 (JDSW) by Lù Démíng 陸德明 (550–630 CE) is

valuable for researchers, especially those interested in Chinese historical phonology, morphonology, and morphosyntax. The creation and sharing of this database is a valuable service to the scholarly community.
The data note is currently deficient in that it is inconsistent and inaccurate in multiple places. The database is currently deficient for similar reasons. Both can be remedied with minimal effort, and should be.

Comments on the Data Note

1. The terms Ancient Chinese, Middle Chinese, Archaic Chinese, and Old Chinese are used inconsistently and inaccurately, causing confusion.

   The terms “Ancient Chinese” and “Archaic Chinese” are no longer used, having been supplanted by “Middle Chinese” and “Old Chinese”, respectively. Because the authors use all four terms without explanation, it is difficult to know what they are intended to refer to. The authors should decide which terms will be used, make sure their use is consistent across the abstract, plain language summary, and main text. The terms should be clearly defined (with dates) in the introduction.

   Additionally, there are some places (e.g. second paragraph of introduction) where “Archaic Chinese” or “Old Chinese” occurs but the authors appear to intend to refer to Middle Chinese. These uses should be double-checked and corrected.

   Finally, it should be noted that the current scholarly consensus is that Old Chinese did not have tones. It is therefore anachronistically inaccurate to speak of “the tones of Old Chinese” or “the tones of Archaic Chinese”. One should rather speak of “Old Chinese precursors of the Middle Chinese tones”, or otherwise clarify the usage.
2. The qù tone (去聲) is mis-translated as “entering tone” in the abstract and plain language summary. It is mis-translated as “falling tone” under General statistics on the data. The translation is missing entirely in the first paragraph of the Introduction. It should be consistently translated as “departing tone”. Moreover, it is inconsistently transcribed as “Qù tone”, “qù tone”, and “qùshēng”.
3. In the first paragraphs, the notations A(1), B(2), C(3), and D(4) for the tones is unclear. The authors should clarify that A, B, C, D and 1, 2, 3, 4 are two alternative notational designations of the four tones of Middle Chinese.
4. It is unclear why *Shuōwén jiězì* 說文解字 is given the date (1st to 2nd c. CE). The scholarly consensus is that the work was completed in 100 CE and presented to the throne in 121 CE. It is therefore appropriate to date it to the 2nd c. CE.
5. The *dúruò* glosses are explained by the incorrect schema X 讀若《丙》. What is “丙” here? The English translation “read [this character] as X” is incorrect. The actual schema found in the dictionary is 讀若 Y, meaning “[The head entry character X] is pronounced like [the character] Y”.
6. The statement that *fanqie* “developed from at least the 3rd c. CE on, based on Hàn period and later glosses to the Classics” is confusing. What does “based on” mean here?
7. “In the *fǎnqiè* system a character is divided into an initial and a final (rime) part (including tones)”. In standard phonological descriptions of Chinese, “final” and “rime” are not equivalent. “Final” includes the pre-nuclear onglide, “rime” does not. In *fanqie*, the second spelling character indicates the final and tone, not just the rime.
8. Table 1 has a column titled “EMC”, but the reader is not told that this stands for “Early Middle Chinese”. Furthermore, Early Middle Chinese was earlier defined as 2nd c. BCE – 3rd

c. CE. But the reconstructions cited in the table are EMC as the term is used by historical phonologists, i.e. they are reconstructions of 6th-century Middle Chinese.

9. Table 1 is confusing in other way. “MM” is not defined. The meaning of the symbols in square brackets after the pinyin spelling is not explained. I cannot figure out what “d [d] [ɔ l]” is supposed to represent. In the third column, the colons look like phonetic symbols but perhaps they are intended to be colons? But why should there be such punctuation in tabular data?

    I suggest redoing this table. The first column should give the character and its English gloss. The second should give its Modern Mandarin pronunciation in pinyin and IPA. The third should give its Middle Chinese reconstructed sound value. The fourth should give the two fanqie spellers. The fifth should give their MM pronunciations in pinyin and IPA. The sixth should give their Middle Chinese reconstructions. The seventh should illustrate how the initial of the first along with the final and tone of the second combine to give the pronunciation of the target character.

    The Middle Chinese system should be indicated and explained. It looks like it might be Pulleyblank, except the “h” for departing tone is not properly superscripted. Since the database itself uses Baxter MC transcriptions, why does Table 1 here use Pulleyblank instead of Baxter?
10. In the paragraphs after Table 1, the authors refer to “cognate sets”. I believe they mean “word families”, as cognate sets would require the comparison of related words across languages. Or perhaps the authors intend to use “cognate” to refer to a word family, i.e. all the words in a single language that share a common root? If so, they should state that clearly.
11. It is not clear what “lexical borrowings” refers to in the quote from Pulini & List, or why such borrowings would play a role in this process.
12. The example, involving the verbs ‘to see’ and ‘to show’, is not clear because only one graphic form is provided for each word. This does not seem to illustrate any of the three issues of alternate graphs, phonetic loans, or lexical borrowing. Are the authors trying to point out that the challenge here is that the root word and its derived word are not written using the same character?
13. “In the present project all *qùshēng* roots involving a verbal reading, and all roots which alternate with a (verbal) cognate (a derivation) in the *qùshēng* are at issue. We include also those cases where a deverbalized noun appears in the *qùshēng*.” This statement is confusing. All verbs in departing tone are included, even if there is no apparent alternation? The use of “cognate” is confusing too — cognate and derivation are not synonyms. Perhaps this can be rephrased in a clearer way.
14. Under “Materials”, the authors should specify the relationship between Unger 1989 and JDSW. Does Unger include all and only the *fanqie* glosses in JDSW? Has the data in Unger been verified or does it contain errors?
15. It is unclear to me from the “Materials” section if the database contains only alternations (i.e. root and derived word) written with the same character, or if it includes example like ‘to see’ and ‘to show’, i.e. a suspected alternation written with two different characters.
16. What does “fǎnqiè of tones A, B, C, D respectively” mean? How can a tone have a *fanqie*?
17. In Table 3, I presume that an “Alternation type” of “1 2 3” means that there are roots and derived forms in the three tones level, rising, and departing. Thus any column that includes “3” indicates a tonal alternation with departing tone. But what do the entries like “1” mean?

Level tone words that have no derivatives in other tones? If so, why are they included at all?

18. “the proportions of alternations with respect to other tones does not show any particular trend in the data”. It is not clear to me what this means. What would consistute a “trend”? In what way do the Qù-tone alternations show a “trend”?
19. In the “Conclusions and Outlook” section, the term “FQ” appears. This is presumably fǎnqiè, and should be expanded as such.
20. “digital dictionary of Archaic Chinese lexical roots in the *qùshēng* (original or derived)”. I don’t understand how a lexical root can be “derived”.

Comments on the database

1. The Pinyin column of the Roots table contains data that is not in a consistent notation. Some entries * before, some don’t. (I can’t find an explanation of this in the text.) Some have tone mark diacritics, some have a tone number following the syllable (e.g. bi4 vs. bì). The notation for multiple pronunciations is also inconsistent, sometimes using a comma, sometimes a slash (with or without spaces), sometimes just tone numbers separate by a comma. A filter on bì bring up three entries (25, 43, 57) but a filter on bi4 brings up 7 different entries (34, 37, 39, 45, 49, 54, 55). All ten should be findable with a single filter.
2. There are also asterisks in the Middle Chinese column, for unclear reasons.
3. Some entries don’t make sense to me. For example, the character 食, #657, has modern pronunciation given as *sì/shí. The MC tone is given as 2 (shǎng), which is not accurate: this character has Middle Chinese qù and rù tones. The last column indicates there is one occurrence, but clicking through to the Occurrences table shows no data.
4. In the Occurrences table, The Middle Chinese reconstruction column contains impossible syllables (showing there is a problem with the algorithm used to generate them). Examples of such impossible syllables are *nryoH* and *nyyoH* and *peajng*.

Is the rationale for creating the dataset(s) clearly described?

Partly

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of methods and materials provided to allow replication by others?

Partly

Are the datasets clearly presented in a useable and accessible format?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I have expertise in Chinese historical phonology and lexicography.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 24 Jan 2026

Johann-Mattis List

Thanks a lot for these very helpful comments. We will carefully revise database and study in the nearer future, trying to account for all points raised here, and then providing also a point-to-point reply.

Competing Interests: No competing interests were disclosed.

Reviewer Report 19 December 2025

https://doi.org/10.21956/openreseurope.22870.r63399

© 2025 Harbsmeier C. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Christoph Harbsmeier
University of Oslo CEES, Oslo, Oslo, Norway

Report on **A database of qù-tone alternations in Ancient Chinese [version 1; peer review: awaiting peer review]**
Barbara Meisterernst1, Johann-Mattis List

This is a disciplined concise paper full of relevant information that is of broad interest to all serious students of classical Chinese linguistics. The publication of this will be helpful. The paper outlines the structure of an informative database that explores highly original dimensions of classical Chinese morphology with a commendable primary source focus.

The available critical edition of the primary source of this paper, the typeset Jingdianshiwen Huang Zhuo, Jingdianshiwenhuijiao, Beijing: Zhonghuashuju, 2006, is not mentioned in the bibliography.

This is worrying because this is the main quite indispensable source that should have been consulted and could have been consulted on crucial evidence.

The reference is instead to Lu D, 德明 陸: Jīngdiǎn Shìwén. 經典釋文[Explanation of the Classics]. Guji Chubanshe: Shanghai, (583[1985]).

The brief introductory material needs rewriting: “Lù Démíngs historical annotation of the classics” needs to be rewritten as something like “collection of phonetic glosses on a selection of classics”

“the formats recommended by the Cross-Linguistic Data Formats initiative. **This** format” is careless.

“access to experienced users” would be clearer as “access for experienced users”.

“The database is available in two flavors” needs to be reformulated.
“might give hints to former grammatical processes in the history of the Chinese language” must be improved.

“The study presents a database in which such alternations have been assembled for alternations involving the so-called entering tone, one of four tones in Old Chinese.” The question whether the “entering tone” is a tone must be addressed.

“The database allows to browse the evidence in various forms.” needs to be rewritten: “allows you to” “allows one to”, one feels.

“The study presents how” needs to be rewritten as “explains how”.

My suggestion is that this paper needs basic language revision.

The authors skilfully integrate various digital resources in their possession, but fail to give an account of the centrepiece of their work, the text of the JDSW arranged as a table or spreadsheet. If they rely on the work of others, they should state so, and if they have inputted and arranged the enormous amount of data themselves, they should state how they have gone about this and which text editions they base themselves on. This would also allow others to utilise their work responsibly, in the spirit of the Cross-Linguistic Data Formats initiative.

The general introduction of this broadly ambitious paper would become more convincing if it contained a much more comprehensive explanatory survey of existing research and a richer bibliography of material directly useful for the study of morphology and derivation by tone change in classical Chinese.

In this paper there is no discussion of source criticism of the phonetic glosses in Jingdianshiwen and and other early sources. Luo Changpei’s (ed. 2008 below) meticulous study of Tang dynasty and later versions of the Jingdianshiwen are indispensable for the crucial textual criticism.

The English bibliography is missing too many obviously relevant publications that one assumes the authors are familiar with. Perhaps the following suggestions might be helpful:
[1] Baxter, W. H., & Sagart, L. 1998. Word formation in old Chinese. *New Approaches to Chinese Word Formation: Morphology, phonology and the lexicon in modern and ancient Chinese*, 35-76.
[1] Dai, J. X. L. 1990. Historical morphologization of syntactic words: Evidence from Chinese derived verbs. *Diachronica*, *7*(1), 9-46.
[1] Jacques, G. 2022. On the nature of morphological alternations in Archaic Chinese and their relevance to morphosyntax. *Bulletin of the School of Oriental and African Studies*, *85*(3), 475–494.
[1] Mazo, O. M. 2002. Reflexes of the suffix*-s in Classical Tibetan and Old Chinese. In *Proceedings of the Ninth Seminar of the IATS, 2000. Volume 1: Tibet, Past and Present* (pp. 455-472). Brill.
[1] Mei, T. L. 2015. Proto-Sino-Tibetan morphology and its modern Chinese correlates. *Oxford handbook of Chinese Linguistics*, 58-67.
[2] Mei, Tsu-Lin. 2012. The causative *s- and nominalizing *-s in Old Chinese and related matters in Proto-Sino-Tibetan. Language and Linguistics, 13(1): 1—28.

[1] Sagart, L. 1999. The origin of Chinese tones. In *Proceedings of the Symposium/Cross-Linguistic Studies of Tonal Phenomena/Tonogenesis, Typology and Related Topics.* (pp. 91-104). Institute for the Study of Languages and Cultures of [1] Schuessler, A. 1976. Affixes in Proto-Chinese. *Mtlnchener Ostasiatische Studien*, Band 18. Wiesbaden: Franz Steiner.
[2] Schuessler, A. 1985. The function of qusheng in Early Zhou Chinese. *Linguistics of the Sino-Tibetan area: the state of the art* ed. by Graham Thurgood, James A. Matisoff and David Bradley, 344—362. Pacific Linguistics series C-No.87. Canberra: The Australian National University.
[1] Schuessler, A. 2009. *Minimal Old Chinese and later Han Chinese: a companion to Grammata serica recensa*. University of hawaii Press.
Harbsmeier, C. 2016 Irrefutable Conjectures. A Review of William H. Baxter and Laurent Sagart, Old Chinese. A New Reconstruction, Monumenta Serica, 64:2, 445-504.
Harbsmeier, C. 2024 "The Cultural Politics of Chinese Lexicography", in Тангутская и китайская филология К юбилею М. В. Софронова, Moscow: Institute of Oriental Studies, 2024, pp. 439-480.
Pulleyblank, Edwin G. 1973. Some Further Evidence Regarding Old Chinese *-s and Its Time of Disappearance. BSOAS 36.2: 368–73.
———. 1991. Some Notes on Morphology and Syntax in Classical Chinese. In Chinese Texts and Philosophical Contexts: Essays Dedicated to Angus C. Graham, ed. J. H. Rosemont. Pp. 21–45. Open Court.
The following Chinese theses are among those directly relevant to refer to:
[1] Du Jifang 2007, Shandong University, 2007.
[2] Gao Yingze, 2011 Tianjin Nankai University, 2011.
[3] Wan Xianchu, Wuhan University, 2002.
[4] Wang Cheng, Beijing Normal University, 2008.
[5] Yang Yifan, Nanjing University, 2011.

The extensive bibliography of relevant Chinese books that are missing in the bibliography is too large to list up.

In sum, I suggest that this paper is substantial and well worth publishing, but that it needs and deserves significant revision before it is published.
The usefulness of this project is dependent on the reliability of its digitised sources, especially that of the text of Jingdian Shiwen. The authors need to clarify how they ensure the integrity of the resources they employ and what are the digitised text editions these are based upon.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
No

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Chinese historical linguistics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 24 Jan 2026

Johann-Mattis List

We are very greatful for these many helpful comments. We will carefully revise both the database and the study presenting it in the nearer future. Once done, we will submit a revised version, trying to account for all open questions raised by the reviewer.

Competing Interests: No competing interests were disclosed.