Towards a standardized annotation of rhyme judgments in Chinese historical phonology (and beyond)

Johann-Mattis List, Nathan W. Hill, and Christopher Forster

Abstract

Although rhyme analysis plays a crucial role in the reconstruction of Old Chinese phonology, the field has not yet developed a standardized annotation framework for rhyme judgments applied to Ancient Chinese texts. Building on initial attempts to standardize cross-linguistic data for the purpose of historical and typological language comparison (as part of the Cross-Linguistic Data Formats initiative), we present a proposal for consistent and transparent rhyme annotation. This proposal allows scholars to annotate the rhymes they identify in historical texts in such a way that the judgments can be analyzed with computational tools as well as conveniently inspected by scholars. Our framework is accompanied by software tools and exemplary datasets, which were annotated by various scholars, and reflect not only Chinese, but also contemporary poetry in different languages. In the paper, we present the framework and also point to caveats and current insufficiencies in annotation. In doing so, we hope to inspire more scholars working on Old Chinese reconstruction to share their judgments, allowing others working in the field to improve, revise, and analyze them.

1 Introduction

Rhyme analysis plays a crucial role for the reconstruction of Old Chinese phonology, but the field has not yet developed a standardized framework for annotating rhyme judgments. In this paper, we want to present a new annotation framework for rhyme judgments, which builds on the general idea of increasing the comparability of data in historical linguistics and language typology, and has the goal of being not only applicable to Chinese texts, but to the poetic traditions of any language that uses rhyme as a device.

In the following, we introduce our framework in detail, by first pointing to the importance of rhyme analysis for Chinese historical phonology (1.1), discussing the typical practice of rhyme annotation in Chinese linguistics (1.2), and presenting some general thoughts on the importance of annotation in philology and linguistics (1.3). We then present our framework in detail, by introducing the Cross-Linguistic Data Formats initiative (2.1), presenting the main ideas for rhyme annotation (2.2), and providing several examples of rhyme annotation in practice (2.3). We conclude by articulating the hope that our example can inspire scholars in our field to improve the transparency of our research by providing data underlying analyses in generally comparable formats.

1.1 Rhyme analysis in Chinese historical phonology

Due to phonetic change, the rhymes of ancient Chinese texts often cease to rhyme in more modern pronunciations. Already in the sixth century of our era Shěn Zhòng 沈重 noticed failures of expected rhymes in the Shījīng 詩經; he suggested adjusting one's pronunciation to make the rhymes read smoothly. The Míng 明 dynasty scholar Chén Dì 陳弟 (1541-1617) explained that sound change had altered the original pronunciation of at least some words, and that these words normally had a single pronunciation in the mouths of the ancients (Baxter 1992, 154). The scholar Gù Yánwǔ 顧炎武 (1613-1682) was first to undertake a reconstruction of the rime categories of Old Chinese; he elaborated ten rime categories (*yùnbù* 韻部) in the *Shījīng*, which split into the more elaborate categories of Middle Chinese rimes (Baxter 1992, 155–57). Subsequent scholars distinguished categories that the *Shījīng* keeps apart in its rhyming practices, which Gù Yánwǔ had failed to notice. The categories recognized by scholars working within the Chinese philological tradition steadily rose over time to 22 (Baxter 1992, 157–71). In the late 20th century, armed with the six vowel hypothesis of Old Chinese, and motivated by the internal reconstruction of Middle Chinese, the three scholars Zhèngzhāng Shàngfāng 鄭張尚芳 (Zhengzhang 2000), Sergei Starostin (Starostin 1989), and William Baxter (Baxter 1992) independently recognized many more rime categories. For example, Schuessler (Schuessler 2009), who also operates in the six-vowel tradition, puts the total number of Old Chinese rime categories at 38 and we count 45 in Baxter & Sagart's most recent Old Chinese reconstruction (Baxter and Sagart 2014).

The rime category of an Old Chinese word is only directly knowable if that word happens to occur as a rhyme word in the *Shījīng* . Except for in those few cases where the Middle Chinese pronunciation of a word may, according to one's overall theory, develop only from a single Old Chinese rime category, in order to speak of the rime category of words that do not appear as rhyme words in the *Shījīng*, one must turn to the phonetic information inherent in the Chinese writing system.

1.2 Rhyme annotation in Chinese historical phonology

The ways in which scholars share their respective rhyme judgments in the literature is very diverse and makes a formal comparison of different rhyme analyses difficult. The problem here lies only to some degree in missing digital versions of important contributions, which would be merely a problem for pure computational approaches). A more significant problem is that many authors report their rhyme judgments in a form that is insufficiently explicit to infer the individual judgments made on individual poems and stanzas. Apart from scholars who presented only the *results* of their analyses, without providing the evidence (郑张尚芳 2003; 潘悟云 2000), we also often find analyses that are extremely difficult to inspect, due to the way they present their judgments. In this sense, only a small amount of rhyme analyses is truly *explicit*.

An example for the problem of insufficient explicitness in the way rhyme judgments are reported is the otherwise excellent study of Old Chinese phonology by Sergej Starostin (Starostin 1989, 458–674): Instead of providing a full version of the *Shījīng* that he used for his reconstruction, Starostin's data starts from rhyme groups and then lists all rhyme words per stanza that he judges to reflect this rhyme group. For example, for the rhyme group *zhī*

之 _*-ə_, we find the rhyme words _*ćə:_ 哉, _*gə_ 其, _*tə_ 之, and _*sə_ 思 (p. 448), which directly corresponds to the classical analysis of stanza 2 in Ode 109, for which Wáng Lì gives the following rhyme judgments:

彼人是哉(tzə)! 子曰何其(giə)!
心之憂矣,其誰知之(tjiə)?
其誰知之(tjiə)?
蓋亦勿思(siə)! (_Shījīng_, 109.2)

Starostin's analysis is more explicit than other attested analyses, in that it makes a formal representation, in which each rhyme word in the text of the _Shījīng_ is marked as such along with the proposed reconstruction. Nonetheless, any attempt to digitize or reverse-engineer individual judgments from the data in the book would require a full digitization and numerous hours of identifying each character's occurrence in the original source. In contrast, Wáng Lì's format is very transparent, insofar as it marks exactly where each rhyme word occurs in context.

Explicit analyses of _Shījīng_ rhymes — apart from Wáng Lì (Wáng 1980) — also include Karlgren (Karlgren 1950), Baxter (Baxter 1992, 583–743), and Wáng Xiǎ'an (Wáng 2011). In all these analyses, the original text of the _Shījīng_ that was taken as the basis for the rhyme judgments is accompanied by a note indicating which lines in each stanza rhyme and how the authors reconstruct the rhyme words in those lines. Here again, however, we can find differences in the degree of explicitness by which authors report their actual rhyme judgments. While Wáng Lì, for example, adopts an annotation that marks rhymes that recur across stanzas, Baxter only shows rhymes inside each stanza. Furthermore, it is rare for any of the authors to point to instances of internal rhyme, probably also due to the fact that their general rhyme annotation schema is built in such a way that it describes the relation between lines in the _Shījīng_ (as opposed to the relation between words inside a stanza or a poem).

1.3 Annotation in linguistics and philology

Annotation is crucial for scientific research involving language and texts. The major idea of annotation is to provide some kind of _added value_ for a given resource (Milà-Garcia 2018), i.e., some information that could not — or only with great efforts — be extracted from the original resource without resorting to intensive search or complex computational algorithms. What value we add when annotating a resource depends on our research question. In _inter-linear-glossed text_ (MPI EVA 2008), for example, linguists try to provide some kind of a meta-language for disentangling grammatical from content words, in order to help other linguists to understand how the general meaning of a phrase or sentence is constructed. In _morphological annotation_, as introduced by Hill and List (Hill and List 2017), the same idea is applied to multi-morphemic forms in cross-linguistic word lists.

One can roughly distinguish two basic types of annotation: *inline* and *stand-off* annotation (Eckart 2012). While inline annotation manipulates the original data directly, for example, by adding tags, stand-off annotation only references the original data, without directly modifying it. Most annotation frameworks, however, typically use a mixture between the two types, although it is clear that stand-off annotation has the advantage of allowing for far more flexibility, especially if adding multiple layers of annotation to a given resource.

As an example illustrating the difference between the two annotation styles, consider the rhyme annotation employed by Baxter (Baxter 1992) as compared to the one by Wáng Lì (Wáng 1980), shown above, for poem 109 (second part of stanza 2 in the Shījīng). While Wáng Lì provides the rhyme judgements inline, Baxter (p. 625) basically uses a stand-off annotation by listing all relevant data in tabular form:

Character	**Pīnyīn**	**MCH**	**OCH**	**Rhyme**
哉	zāi	tsoj	*tsɨ	B
其	jī	ki	*k(r)jɨ	B
之	zhī	tsyi	*tjɨ	B
之	zhī	tsyi	*tjɨ	B
思	sī	si	*sjɨ	B

Table 1: Rhyme annotation in Baxter (1992), Ode 109, Stanza 2.

Both types of annotation have advantages and disadvantages. Wáng presents the whole text, so we know exactly which words he judges to rhyme and where he locates the relevant rhyme words. Since Baxter does not provide an index to the words in the original *Shījīng* text, we cannot know exactly where the rhyme words occur in the lines (it is, for example, possible that a character is repeated throughout the same line), and we can also not see the poem as a whole, along with its structure of rhyming and non-rhyming lines. The advantage of Baxter's system, however, is that it allows him to list more data related to each word, including the Pīnyīn transliteration, Middle Chinese and Old Chinese readings, and even his assessments as to which lines rhyme with each other. Thus, while Baxter loses explicitness with respect to the underlying *Shījīng* text, Wáng loses the flexibility of annotation. Ideally, an advanced annotation framework for rhyme judgments should allow for the advantages of both approaches.

2 Towards a standard of rhyme annotation in Chinese historical phonology (and beyond)

As we have seen in the foregoing discussion, the annotation of rhymes — be it in Chinese historical phonology or in general — is not trivial, in particular since there are considerable desiderata for common rhyme annotation frameworks. Thus, we would first like to be able to annotate large collections of poems, like the *Shījīng*, where we retain the original text, but could also indicate character readings, as proposed by different authors in the literature. We may also want to indicate details of rhyming, for example, pointing to impure rhymes or indicating internal rhymes, which we know occasionally occur in the *Shījīng*.

In order to advance our understanding of rhyming in China, we will in the long run require a more comparative, typolological perspective that could tell us to which degree the rhyme practice that we observe in ancient Chinese texts is peculiar or expected. For this reason, it would also be desirable if our rhyme annotation framework could be used for all kinds of rhyming poetry, stemming from different genres, languages, and cultures. Judging from our knowledge of different genres, both in the history of Chinese poetry, but also of poetry world-wide, we may occasionally want to add a lot more information, for example on meter, syllables, word boundaries, or tonal patterns.

While all these aspects need to be taken into consideration when proposing a first format for rhyme annotation, it is also important to be pragmatic to some degree, since we know from experience that very complex format prescriptions will scare off users rather than encouraging them to take part. Finding the right balance between pragmatism and perfectionism is thus crucial for our endeavor.

2.1 The Cross-Linguistic Data Formats initiative

The Cross-Linguistic Data Formats initiative (https://cldf.clld.org) is an attempt to standardize different types of data which are frequently used in the context of historical linguistics and linguistic typology (Forkel et al. 2018). While the current version mainly focuses on standardized formats for wordlists and structural data, the specifications are intended to be expandable in future versions, and draft proposals for dictionaries and parallel texts are underway.

The common procedure of adding new format specifications to the CLDF initiative is by testing the ideas on sufficiently large amounts of data first, before an official discussion of whether and how to integrate a new data format into the CLDF framework should be undertaken. The attempts described here are a first effort at presenting our basic ideas to a broader public, in the hope that after sufficient testing and discussion we can include rhyme annotation frameworks in future versions of the CLDF. Although rhyme analyses of the depth as we propose here are — at least to our knowledge — a rather new enterprise, we are confident that our format proposals are sufficiently useful for inclusion in the CLDF initiative, because they would allow focus on new, fascinating, and largely unexplored cross-linguistic data.

2.2 Main ideas for rhyme annotation

The main ideas for our proposed format of rhyme annotation follow largely the ideas that drove the development of the CLDF format, and although our current proposal has to be seen as independent of CLDF, we hope that the ideas can later be included into a new release of CLDF that would include poems and rhyme annotations as an additional component. The major criteria for the choice of our format proposal follow to a large degree the — among programmers well-known — "Zen of Python", which claims that "Simple things should be simple, complex things should be possible".

Our basic ideas thus require: (1) simplicity, (2) exhaustiveness, (3) flexibility. Simplicity means that people should be able to apply our format prescriptions with a minimal amount of work, using standard off-the-shelf tools, like text or spreadsheet editors, rather than complex new tools that would have to be created specifically for rhyme analysis. Exhaustiveness means that we wish to be able to reflect all knowledge that can be formalized in a given rhyme analysis. While we would always allow adding ad-hoc information in note-fields, we want to offer a high degree of granularity in annotations, allowing, for example, the inclusion of phonetic transcriptions and phonetic alignments (List 2014). Flexibility allows for a quick extension of the data when needed, using mechanisms already offered by the framework.

In order to achieve all these goals, we draw largely from our experience with the enhanced annotation and computer-assisted manipulation of *wordlists* in historical linguistics (Hill and List 2017) and their subsequent inclusion into the CLDF specifications.

2.2.1 Representing rhyme collections in spreadsheets

Following the basic idea of CLDF to represent most of the data in the form of spreadsheets, we propose a very straightforward way to represent rhyme annotations in spreadsheet format. While CLDF proper would require that the data is delivered in form of comma-separated or tab-separated value (CSV or TSV), the data can be easily annotated with widely used spreadsheet editors, such as Excel or LibreOffice. The key component of a spreadsheet is a header line that indicates the values that we find in the sheet, and the rows, that add values for each column as it is described by the header.

Based on the discussions of the desiderata and past experiments which proved the particular insufficiency of certain annotation forms, our core annotation of a poem or a poem collection now contains the following main components:

- **ID:** the identifier, which is a numerical ID.
- **POEM:** a name for the given poem.
- **STANZA:** the stanza of the poem (usually a numeric value, preceded by the name of the poem).
- **LINE_IN_SOURCE:** the line of the poem as we find it in the source from which the data is taken (especially containing original punctuation etc.).
- **LINE:** a double-segmented version of the line, in which words are separated with help of + as a separator, and spaces can be used to represent segments of phonetic values (similar to the format adopted by the LingPy software package to represent phonetic sequences and alignments).
- **LINE_ORDER:** A numerical value that provides the order of the lines of a poem in a given stanza.

- **RHYMEIDS:** A list of numerical identifiers, indicating which words in a the LINE rhyme by assigning the same ID to different words, using 0 to indicate that a given word does not rhyme.
- **ALIGNMENT:** A double-segmented version of the line that can, however, store aligned content, differing from the data in LINE, as well. This data comes in handy when trying to check questions of phonetic similarity of rhyme words, or of vowel purity, which would greatly facilitate automatic analyses as the one presented in List et al. (2017).

With these eight columns provided, poems can be annotated in a very straightforward way, regardless of the language in which they were written. One can, of course, add many more columns, depending on specific characteristics of the datasets, but for the general rhyme annotation, we think that these fields will be sufficient for most of the cases; it substantially exceeds rhyme annotation frameworks that have been proposed so far in terms of detail.

As an example, consider (again) ode 109, stanza 2, in the rhyme judgments of Wáng Lì (Wáng 1980), shown in the table below. Note that the entry for LINE_IN_SOURCE is not shown in Table 2, as the excess length of each row would run beyond the width of this paper, thereby disorienting readers; however it is still a crucial component for this annotation standard, and readers can see the full analysis by Wáng Lì in the supplementary data accompanying this paper.

ID	**POEM**	**ST.**	**LO**	**LINE**	**ALIGNMENT**	**RHYMEIDS**
1733	園有桃	109.2	1	園 + 有 + 棘	園 + 有 + kiək	0 0 467
1734	園有桃	109.2	2	其 + 實 + 之 + 食	其 + 實 + 之 + djiək	0 0 0 467
1735	園有桃	109.2	3	心 + 之 + 憂 + 矣	心 + 之 + 憂 + 矣	0 0 0 0
1736	園有桃	109.2	4	聊 + 以 + 行 + 國	聊 + 以 + 行 + kuək	0 0 0 467
1737	園有桃	109.2	5	不 + 我 + 知 + 者	不 + 我 + 知 + 者	0 0 0 0
1738	園有桃	109.2	6	謂 + 我 + 士 + 也 + 罔 + 極	謂 + 我 + 士 + 也 + 罔 + qiək	0 0 0 0 0 467
1739	園有桃	109.2	7	彼 + 人 + 是 + 哉	彼 + 人 + 是 + tzə	0 0 0 468
1740	園有桃	109.2	8	子 + 曰 + 何 + 其	子 + 曰 + 何 + giə	0 0 0 468
1741	園有桃	109.2	10	其 + 誰 + 知 + 之	其 + 誰 + 知 + tjiə	0 0 0 468
1742	園有桃	109.2	10	其 + 誰 + 知 + 之	其 + 誰 + 知 + tjiə	0 0 0 468
1744	園有桃	109.2	12	蓋 + 亦 + 勿 + 思	蓋 + 亦 + 勿 + siə	0 0 0 468

Table 2: Rhyme annotation format (excerpt) with alignments and identifiers for rhyme words.

While this representation may look complicated at first, it offers a degree of explicitness we have not found in any of the transparent rhyme annotations proposed in the past. On the one hand, we manage to avoid a complex inline annotation, while on the other hand we can express in a very detailed way which words (or characters) in the stanza rhyme, and how they should be pronounced.

In addition, the ALIGNMENT column allows us an even greater detail of the representation of our rhyme analysis, since we can use the column to share explicit phonetic alignments of our data, allowing for a much more fine-grained analysis of questions regarding impure rhymes.

ID	ALIGNMENT	RHYMEIDS
1733	(k) i ə k	467
1734	(d^j) i ə k	467
1735		
1736	(k^w) - ə k	467
1737		
1738	(q) i ə k	467
1739	(tz) - ə	468
1740	(g) i ə	468
1741	(t^j) i ə	468
1742	(t^j) i ə	468
1744	(s) i ə	468

Table 3: Illustrating the power of alignments in our rhyme annotation format.

Comparing this new format proposal with previous annotation frameworks, we can easily see that the possibility of annotating the similarity of rhyme words in the form of *phonetic alignments* offers a multitude of future possibilities, especially when more datasets are annotated in this form. Alignments would allow us not only to access automatically or formally the similarity between two or more rhyme words, they would also allow us to investigate cases of impure rhyming on a large scale, drawing statistics not only across poems that appeared in different epochs of the same language, but also across languages and cultures.

2.2.2 Software API for curation and analysis of rhyme datasets

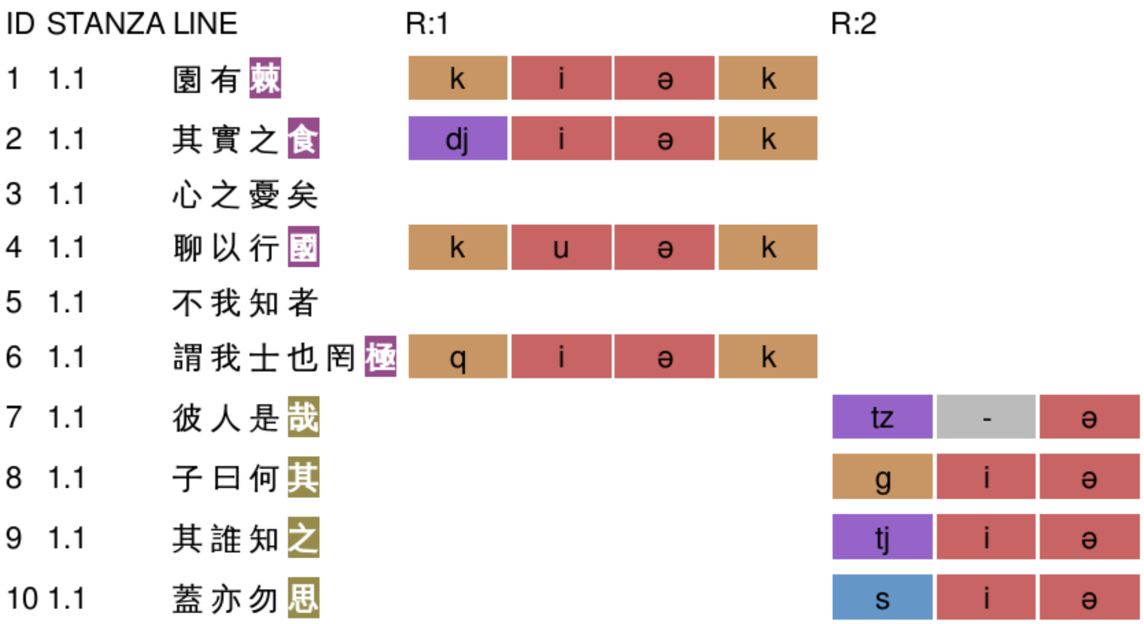
We have developed a software API, called PoePy (https://github.com/lingpy/poepy), that allows one to parse, manipulate, and convert files following our new rhyme annotation schema in a convenient way, with help of the Python language. The framework builds heavily on LingPy, a Python library for quantitative tasks in historical linguistics (List, Greenhill, and Forkel 2017), as well as SinoPy, a Python library for specialized tasks in Chinese historical linguistics (List 2018b). The GitHub site of our API offers additional information for installing and using our software library.

PoePy can read datasets in our general format mentioned above, it can also be used to align rhyme words, provided they are readily assigned to the data, and it can convert the data to different formats, that ease rhyme pattern inspection. Our stanza 2 from Ode 109 of the *Shījīng*, for example, can be rendered directly in the following tabular form, that greatly facilitates seeing the rhyme structure of the poem.

ID	STANZA	LINE	R:467	R:468
1733	109.2	園有**棘**	kiək	
1734	109.2	其實之**食**	djiək	
1735	109.2	心之憂矣		
1736	109.2	聊以行**國**	kuək	
1737	109.2	不我知者		
1738	109.2	謂我士也罔**極**	qiək	
1739	109.2	彼人是**哉**		tzə
1740	109.2	子曰何**其**		giə
1742	109.2	其誰知**之**		tjiə
1744	109.2	蓋亦勿**思**		siə

Table 4: Tabular representation of the rhyme schema underlying stanza 2 in Ode 109.

PoePy can also be used to output the data to HTML format, which allows for a convenient color-coding of rhyme patterns. This format can both be useful for inspection of datasets, or for sharing annotated rhyme data online. An example for our stanza 2 from Ode 109 from the *Shījīng* is given in Figure 1 below.



ID	STANZA	LINE	R:1				R:2		
1	1.1	園有棘	k	i	ə	k			
2	1.1	其實之食	dj	i	ə	k			
3	1.1	心之憂矣							
4	1.1	聊以行國	k	u	ə	k			
5	1.1	不我知者							
6	1.1	謂我士也罔極	q	i	ə	k			
7	1.1	彼人是哉					tz	-	ə
8	1.1	子曰何其					g	i	ə
9	1.1	其誰知之					tj	i	ə
10	1.1	蓋亦勿思					s	i	ə

Figure 1: Colored HTML-output. Colors of the alignments in Wáng Lì's reconstruction indicate the basic sound class to which the sounds belong (alveolars, affricates and velars, vowels).

Given that our current format is rather tedious to produce, PoePy also offers a convenient parser from a much simpler format specification that uses inline-annotation of rhymes. In this format, the same Ode 109, stanza 2, would be rendered as follows:

```
@title: Ode 109
@annotator: Wáng Lì

園有[a/kiək]棘
其實之[a/djiək]食
心之憂矣
聊以行[a/kuək]國
不我知者
謂我士也罔[a/qiək]極
彼人是[b/tzə]哉
子曰何[b/giə]其
其誰知[b/tjiə]之
蓋亦勿[b/siə]思
```
Example 1: Inline format for Wáng Lì’s analysis of Ode 109, Stanza 2.

Thus, one can see that the annotation can be easily achieved by using minimal inline markup, namely square brackets to indicate the rhyme (which is represented by alphabet letters here), along with the option to mark the reading. In a similar way, this format can also be used for a quick annotation of poetry in general. As an example, consider the following excerpt from Mike Naumenko's song "Leto, Pesnja dlja Tsoja" (*Summer, a song for Tsoj*, 1982).

```
@title: Leto. Pesnja dlja Tsoja
@author: Mike Naumenko
@year: 1982
@publisher: ËRIO
@collection: LV
@editor: Mike Naumenko

[a]Лето!
Я изжарен, как кот[a]лета.
Время есть, а денег нету,
Но мне на это напле[b]вать.

[a]Лето!
Я купил себе га[c]зету.
Газета есть, а пива [c]нету.
И я иду его ис[b]кать.
```
Example 2: Inline format for Mike Naumenko’s song *Leto* (“summer”)

The first line is used to store the metadata, which is provided as a pair of a keyword and a value, while the following lines list the poem, separating different stanzas by adding a blank line. Once loading this file in text format with the PoePy library, the data can again be directly queried by printing a table illustrating the rhyme structure, or by querying general statistics about the data. These statistics would, for example, tell us that the song has 119 words in total, 32 lines, 8 stanzas, and 29 rhyme words. From this raw text form based on inline annotation, the data can, of course, also be directly converted to our more refined and flexible format, from where it can be further annotated.

ID	STANZA	LINE	R:1	R:2	R:3
1	1.1	*Лето!*	лето		
2	1.1	Я изжарен, как *котлета.*	кот лета		
3	1.1	Время есть, а денег нету,			
4	1.1	Но мне на это *наплевать.*		напле вать	
5	1.2	*Лето!*	лето		
6	1.2	Я купил себе *газету.*			га зету
7	1.2	Газета есть, а пива *нету.*			нету
8	1.2	И я иду его *искать.*		ис кать	

Table 5: The first two stanzas of the song *Leto*. Since rhyme markers were placed in the middle of the rhyming words, they are now used to split the words into rhyming and non-rhyming parts.

2.3 Examples

2.3.1 Sample datasets

We have started to collect a number of sample datasets that we use for the illustration of our new formats. The largest collection includes the rhyme judgments by Baxter (1992) and Wáng (1980) for the *Shījīng*. In addition, we have started to annotate many small pieces of literature, especially poems, but also popular songs in different languages, which we use to illustrate the usefulness of our annotation system. In the future, we hope to be able to add more datasets in a more consistent manner, digitizing specifically alternative rhyme judgments of the *Shījīng* (such as the those of Karlgren 1950 and Starostin 1989), but also less frequently analyzed rhyme collections, especially from Hàn times.

2.3.2 Rhymes across languages and genres

In the following, we quickly illustrate how our format can be used to annotate rhymes in a much more consistent way than has been done before. Our collection is not bound to a particular language or a particular culture. On the contrary, since the goal of our annotation framework is to provide a much more profound way of annotating formed speech, we have tried to illustrate its usefulness by collecting small examples from different languages and genres.

As a first example, consider Joseph von Eichendorff's Eichendorff (1788-1857) poem *Zwielicht*, which was published as part of a novel in 1815. This poem contains four stanzas of four lines each, all written in form of an "envelope rhyme" (with the general schema "abba"). Our annotation example of stanza 1.1, in which we render the rhyme words in IPA and align them, putting non-rhyming parts of the words in brackets, makes it easy to quickly identify the impure rhyming of the first and the fourth line, which reflects a general peculiarity of German rhyming, in that the diphthongs [ai] and [ɔi] can rhyme freely.

ID	ST	LINE	R:1	R:2
1	1.1	Dämmrung will die Flügel *spreiten*	(ʃ p - r) ai t ə n	
2	1.1	Schaurig rühren sich die *Bäume*		(- b) ɔi m ə
3	1.1	Wolken ziehn wie schwere *Träume* -		(t r) ɔi m ə
4	1.1	Was will dieses Graun *bedeuten?*	(- b ə d) ɔi t ə n	

Table 6: Eichendorff's *Zwielicht* (first stanza) in aligned form.

As another example, consider the first stanza of Bob Dylan's song "I want you" (from the album *Blonde on Blonde*, 1966). Here the rhyme patterns are more complex than in Eichendorff's poem, but rhyming is in parts also more lax, with more imperfect rhymes, reflecting the typical style of Dylan's poetry.

ID	ST	LINE	R:1	R:2	R:3
1	1.1	The guilty undertaker *sighs*	s - ai s		
2	1.1	The lonesome organ grinder *cries*	k r ai s		
3	1.1	The silver saxophones *say*	s - æi -		
4	1.1	I should *refuse_you*		r i f j uː s j uː	
5	1.1	The cracked bells and washed-out *horns*			h - ɔ r n s
6	1.1	Blow into my face with *scorn,*			s k ɔ r n -
7	1.1	but it's not that way, I wasn't *born*			b - ɔ r n -
8	1.1	to *lose* you		- - - l uː s j uː	

Table 7: Bob Dylan's *I want you* in aligned form.

As a further example, the following table presents the first and the third stanza from the famous Chinese song "Yuèliàng dàibiǎo wǒ de xīn", which was popularized in the 1977 version by Teresa Teng. In our analysis of this song, lines 5 and 12 are believed to rhyme with rhyme group R:1, which may be problematic, as it seems that not all native speakers of Mandarin Chinese accept rhymes of *-en* [ən] and *-in* [in]. However, since our analysis will make the overall rhyme schema of the song appear much more harmonic, we think that this reflects the intention of the song writer.

ID	ST	LINE	R:1	R:3
1	1.1	你問我愛你有多**深**	sh ēn	
2	1.1	我愛你有幾**分**	f ēn	
3	1.1	我的情也**真**	zh ēn	
4	1.1	我的愛也**真**	zh ēn	
5	1.1	月亮代表我的**心**	x īn	
11	1.3	輕輕的一個**吻**	w ěn	
12	1.3	已經打動我的**心**	x īn	
13	1.3	深深的一段**情**		q íng
14	1.3	叫我思念到如**今**		l ìng

Table 8: Rhyme annotation for *The moon expresses my heart*

This case shows that the question of whether a given rhyme is indeed intended by a poet or not, may not always be easily solved, and precisely for this reason it is necessary to have frameworks in which the analyses of different readers can be compared. A further example is the song *Te doy una canción* by Silvio Rodriguez (from the album *Mujeres*, 1978), in which none of the three rhyme pairs which we have annotated in stanza 1.2 rhymes perfectly. One might thus assume that rhyming was generally not intended in this song, but we find a very similar pattern in stanza 1.4., and songs in which the words *tú* "you" and *luz* "light" co-occur in potential rhyming position are very frequent in Spanish songs. Our hope is, that with a growing body of datasets in this form, we may learn more about the difference between rhymes which are intended and rhymes which might occur simply by chance.

ID	ST	LINE	R:1	R:2	R:3
7	1.2	Te doy una canción si abro una *puerta*	puer ta		
8	1.2	Y de las sombras sales *tú*		tú	
9	1.2	Te doy una canción de *madrugada,*	madruga da		
10	1.2	Cuando más quiero tu *luz*		luz	
11	1.2	Te doy una canción cuando apareces			
12	1.2	El misterio del *amor*			a mor
13	1.2	Y si no apareces, no me importa:			
14	1.2	Yo te doy una *canción*			can ción

Table 9: Silvio Rodriguez' "Te doy una canción": are the rhymes intended?

As a final example in this section, let us get back to rhyming of texts in Classical Chinese. In Weingarten (2016), rhyming passages in the work of Confucius are presented and analyzed. Given that such examples might provide very valuable evidence for the reconstruction of Old Chinese phonology (and its development thereafter), it would be desirable if a general corpus could be constructed in which all pieces of evidence that can be found throughout different epochs of Chinese language history could be assembled. If we compare the original annotation provided in the text by Weingarten with our extended schema, we think it is obvious how much standardized representations of rhyme judgments, collected collaboratively by all experts in the field, could advance our knowledge about the history of Chinese phonology.

ID	ST	LINE	R:1	R:2
1	1	夫人君無諫臣則失**政**	t e ŋ h	
2	1	士無教交則失**德**	lh ê ŋ h	
3	1	狂馬不釋其**策**		tsh r ê k
4	1	操弓不返於**檠**	ɡ r e ŋ	
5	1	木受繩則**直**		d r ə k
6	1	人受諫則**聖**	lh e ŋ h	
7	1	受學重問		
8	1	孰不順**成**	d e ŋ	
9	1	毀仁惡士		
10	1	且近於**刑**	ɡ ê ŋ	
11	1	君子不可以不學		

Table 10: Rhymes in Confucius' work (as detected by Weingarten 2016).

2.3.4 Comparing differences in rhyme annotations

In List et al. (List et al. 2017), rhyme networks were used to test to which degree different reconstruction systems conform to what Ho (2016) calls "vowel purity", namely the hypothesis that rhyming practice in Old Chinese (and probably also later) was very strict in adhering to identical vowels in rhyming. The test by List et al. (2017) revealed that the system of Baxter and Sagart (2014) (and of six-vowel theories of Old Chinese in general) reflects the principle of vowel purity much more closely than do systems with more vowels (Karlgren 1950) or fewer vowels (Wáng 1980; Lǐ 1971).

In this context, it is important to recall that — what was also mentioned in the paper by List et al. (2017), but might easily be misunderstood by readers — the adherence to vowel purity cannot be used to prove or disprove a given reconstruction system, since the adherence to vowel purity is a hypothesis about Old Chinese rhyming practice itself, and we know well that vowel purity in rhyming can be easily abandoned or disregarded across rhyming traditions in different cultures. Apart from the problem that studies on vowel purity do not bear any diagnostic value with respect to the accuracy of reconstruction systems, one additional problem in the study by List, et al. (2017) is the fact that vowel purity itself was only tested by comparing the rhyme judgments of one source Baxter

(1992) with different reconstruction systems. Given that Baxter himself is reconstructing a six-vowel system on the basis of rhyme evidence, it is quite likely that the rhyme decisions proposed by Baxter (1992) could have influenced the analysis.

While alternative rhyme judgments were not available when drafting the original study on vowel purity, we have now, thanks to our new format for rhyme annotations, also had the time to digitize the rhyme judgments reported in 王力 (1980). Given that two different rhyme analyses have been digitized now, it is interesting and also important for the reconstruction of Old Chinese Phonology to check to which degree different scholars differ in what they judge to rhyme and what not.

We can think of different measures to compare the difference in the actual rhyme judgments of the two versions. A simple measure is to compare how many stanzas differ. From 1070 common stanzas, 175 are different between Wáng and Baxter, which amounts to 15.9%. A far more interesting aspect is to check *how much* different stanzas differ. Similar to a common partitioning task by which we compare to which degree two partitionings of the same data differ, we can do this with help of the B-Cubed scores (Amigó et al. 2009; List, Greenhill, and Gray 2017), since the assessment for a given stanza, whether two words rhyme or not, can also be thought of as a clustering task (authors decide which words belong to the same rhyme partition in a given cluster). Applying B-Cubed scores to compare the rhyme judgments, with help of the PoePy library, to which we added a function to compare different rhyme judgments (implementing the code presented in List 2018a), we find 97% of similarity between Baxter's and Wáng's rhyme judgments. This means that the internal difference between the rhyme judgments by Baxter and Wáng is less pronounced than one might think when only checking whether a given stanza is interpreted differently in *any* way.

ID	**ST**	**LINE**		**R:331**
1208	71.1	綿綿葛藟		
1209	71.1	在河之**滸**		xa
1210	71.1	終遠兄弟		
1211	71.1	謂他人**父**		biua
1213	71.1	亦莫我**顧**		ka

(a) Wáng's rhyme analysis.

ID	**ST**	**LINE**	**R:319**	**R:320**
1229	71.1	綿綿葛**藟**	藟	
1230	71.1	在河之**滸**		滸
1231	71.1	終遠兄**弟**	弟	
1232	71.1	謂他人**父**		父

ID	ST	LINE	R:319	R:320
1234	71.1	亦莫我顧		顧

(b) Baxter’s analysis

Table 11: Comparing Wáng’s and Baxter’s rhyme analysis of Ode 71, Stanza 1. For Baxter’s analysis, our current digitized version does not have the original reconstructions, which is why the software only shows the rhyming characters instead.

As an example for differences in Baxter's and Wáng's rhyme annotations, compare stanza 1 in Ode 71, which is given in the version of both authors below. As can be seen from this example, both authors agree regarding the rhyming of *xǔ* 滸, *fù* 父, and *gù* 顧, but while in Wáng's analyses these three characters are the only ones that write rhyming words, Baxter's analysis assumes in addition, that *lěi* 藟 and *dì* 弟 rhyme as well.

3 Summary and Outlook

In this paper, we have proposed a new framework for rhyme annotation that can be used for a more consistent rendering of the rhyme judgments proposed by different scholars. The framework is inspired by general attempts to standardize cross-linguistic data within the Cross-Linguistic Data Formats initiative, and offers a software library that can be used to check, curate, and analyze rhyme data which has been annotated according to our format specifications. We have illustrated the usefulness of the framework by providing examples of how different cases can be handled. Thanks to the format, we can furthermore easily compare different rhyme annotations in a consistent way. In the future, we hope to expand the so far rather small database of rhyme annotations we have assembled so far. We hope, however, also that our annotation framework will convince our fellow colleagues to help increase the evidence for Old Chinese reconstruction by publishing their future rhyme analyses in a transparent form. Given the multitude of open problems related to the history of the Chinese language from its origins until today, we will only be able to advance our field when working in collaboration and sharing our data in a transparent form.

Source Code and Data

The data discussed in this paper is available along with the PoePy library, which can be accessed on GitHub at https://github.com/lingpy/poepy, and will be officially released in case this paper gets accepted. The code to run the experiments discussed in this paper (especially the comparison of two rhyme datasets) is also available from this repository. A full tutorial introducing the most important aspects of the library is currently being planned.

References

Amigó, Enrique, Julio Gonzalo, Javier Artiles, and Felisa Verdejo. 2009. “A Comparison of Extrinsic Clustering Evaluation Metrics Based on Formal Constraints.” *Information Retrieval* 12 (4). Hingham, MA, USA: Kluwer Academic Publishers: 461–86.

Baxter, William H. 1992. *A Handbook of Old Chinese Phonology*. Berlin: de Gruyter.

Baxter, William H., and Laurent Sagart. 2014. *Old Chinese. A New Reconstruction*. Oxford: Oxford

University Press. http://ocbaxtersagart.lsait.lsa.umich.edu/BaxterSagartOC2015-10-13.xlsx.

Eckart, Kerstin. 2012. “Resource Annotations.” In *CLARIN-d User Guide*, edited by AP 5 Clarin-D, 30–42. Berlin: DWDS.

Forkel, Robert, Johann-Mattis List, Simon J. Greenhill, Christoph Rzymski, Sebastian Bank, Michael Cysouw, Harald Hammarström, Martin Haspelmath, Gereon A. Kaiping, and Russell D. Gray. 2018. “Cross-Linguistic Data Formats, Advancing Data Sharing and Re-Use in Comparative Linguistics.” *Scientific Data* 5 (180205): 1–10.

Hill, Nathan W., and Johann-Mattis List. 2017. “Challenges of Annotation and Analysis in Computer-Assisted Language Comparison: A Case Study on Burmish Languages.” *Yearbook of the Poznań Linguistic Meeting* 3 (1): 47–76.

Ho, Dah-an. 2016. “Such Errors Could Have Been Avoided. Review of “Old Chinese: A New Reconstruction". by William H. Baxter and Laurent Sagart.” *Journal of Chinese Linguistics* 44 (1): 175–230.

Karlgren, Bernhard. 1950. *The Book of Odes: Chinese Text, Transcription and Translation*. Stockholm: Museum of Far Eastern Antiquities.

List, Johann-Mattis. 2014. *Sequence Comparison in Historical Linguistics*. Düsseldorf: Düsseldorf University Press.

———. 2018a. “More on Network Approaches in Historical Chinese Phonology (音韵学).” In *The 2nd Li Fang-Kuei Society Young Scholars Symposium*, 157–74. Taipei: Li Fang-Kuei Society for Chinese Linguistics.

———. 2018b. “SinoPy: A Python Library for Quantitative Tasks in Chinese Historical Linguistics.” Jena: Max Planck Institute for the Science of Human History. https://github.com/lingpy/sinopy.

List, Johann-Mattis, Simon J. Greenhill, and Russell D. Gray. 2017. “The Potential of Automatic Word Comparison for Historical Linguistics.” *PLOS ONE* 12 (1). Public Library of Science: 1–18.

List, Johann-Mattis, Simon Greenhill, and Robert Forkel. 2017. *LingPy. A Python Library for Quantitative Tasks in Historical Linguistics* (version 2.6). Jena: Max Planck Institute for the Science of Human History. doi:https://doi.org/10.5281/zenodo.1065403.

List, Johann-Mattis, Jananan Sylvestre Pathmanathan, Nathan W. Hill, Eric Bapteste, and Philippe Lopez. 2017. “Vowel Purity and Rhyme Evidence in Old Chinese Reconstruction.” *Lingua Sinica* 3 (1): 1–17.

Milà-Garcia, Alba. 2018. “Pragmatic Annotation for a Multi-Layered Analysis of Speech Acts: A Methodological Proposal.” *Corpus Pragmatics* 2 (1): 265–87.

MPI EVA. 2008. *The Leipzig Glossing Rules: Conventions for Interlinear Morpheme-by-Morpheme Glosses*. Leipzig: Max Planck Institute for the Science of Human History. http://www.eva.mpg.de/lingua/resources/glossing-rules.php.

Schuessler, Axel. 2009. *Minimal Old Chinese and Later Han Chinese. A Companion to Grammata Serica*. Honolulu: University of Hawai’i Press.

Starostin, Sergej Anatol’evič. 1989. *Rekonstrukcija Drevnekitajskoj Fonologičeskoj Sistemy (Reconstruction of the Phonological System of Old Chinese)*. Moscow: Nauka.

Wáng, Xiǎ 王显. 2011. *Shījīng Yùnpǔ 诗经韵谱*. Běijīng: Shāngwù 商務.

Weingarten, Oliver. 2016. “The Singing Sage: Rhymes in Confucius Dialogues.” *Bulletin of SOAS* 79 (3).

Zhengzhang, Shangfang. 2000. *The Phonological System of Old Chinese*. Translated by Laurent Sagart. Paris: École des Hautes Études en Sciences Sociales.

李方桂, Li Fang-kuei. 1971. “Shànggǔyīn Yánjiū.” *Qīnghuá Xuébào 清華學報* 9 (1-2): 1–60.

潘悟云, Pān Wùyún. 2000. *Hànyǔ Lìshǐ Yīnyùnxué*. Shànghǎi 上海: Shànghǎi Jiàoyù 上海教育.

王力, Wáng Lì. 1980. *Shījīng Yùndú*. Shànghǎi 上海: Shànghǎi Gǔjī 上海古籍.

郑张尚芳, Zhèngzhāng Shàngfāng. 2003. *Shànggǔ Yīnxì*. Shànghǎi 上海: Shànghǎi Jiàoyù 上海教育.