**Using Python Tkinter to Prototype a Chinese Phonology Database**

**Zhaozhong Zhu**

**Advisor: Dr. Fang-yi Chao**

**Abstract**

Existing research in Chinese phonology has been limited in scope to simply identifying and understanding historical pronunciation tables. However, how these tables have evolved over time and how these changes reflect the linguistic evolution of the Chinese language has not yet been fully explored. Using Python’s Tkinter module, the standard Python Graphical User Interface (GUI), this project seeks to resolve this shortcoming by introducing a prototype that can display the phonological change of a character either by time or by a specific scholar’s interpretation. This application provides researchers with a lightweight tool that seeks to link separate phonological tables into a single record.

**Introduction**

In the English language, words are formed by the 26-character alphabet, and English speakers derive pronunciation from the phonetic rules that govern the sounds of consonants and vowels. However, no such analogy can be drawn to the Chinese language, in which mono-syllabic characters, individually or collectively, form a phrase with a corresponding meaning. Put simply, the characters that serve as building blocks for English words simply do not exist in Chinese. Thus, before the rather recent introduction of *pinyin*, a different methodology existed to codify pronunciation of thousands upon thousands of these characters. A brief overview of these specific “pronunciation dictionaries” or phonology tables are mentioned in the below section.

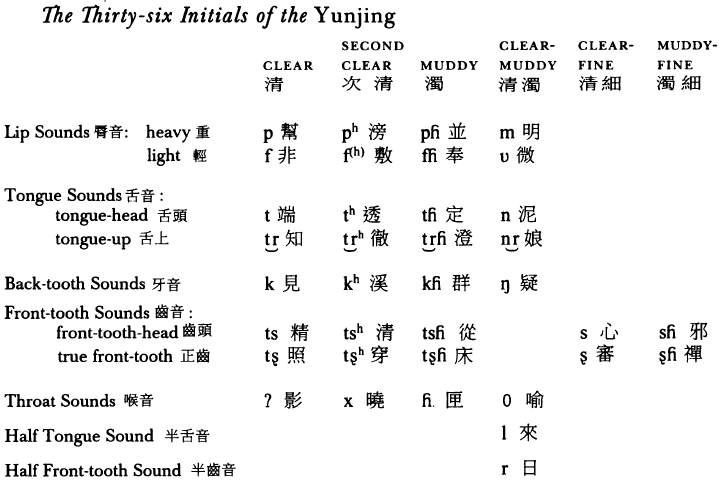
While we do not seek to explore the specific details surrounding the intricacies of Chinese linguistics and historical phonology, we review relevant background information that serves as a backdrop for our research question. Therefore, the following sections are intended to give a high level overview to readers that maybe unfamiliar with central documents surrounding historical phonology.

This paper is not intended to be interpreted as a strictly technical paper or a strictly linguistic one. Instead, we explore the application of existing software to Chinese phonology, and its application to current and future research in this particular field.

**Literature Review and Overview**

Edwin Pulleyblank identifies two primary sources as the foundation for Chinese historical phonology. The first, the *Qieyun* 切韵 (Early Middle Chinese, A.D. 601), “represents the codification of an elite standard going back ultimately to Luoyang in the third century but surviving during the period of division in two gradually diverging forms in north and south China.”[[1]](#footnote-1) The dictionary, however, was not an entirely new invention; rather, it served as a compilation of five previous dictionaries, four from the north and one from the south. One of the most noteworthy achievements of this dictionary is the method in which pronunciation was recorded, namely through the *fanqie* method. Due to the mono-syllabic nature of Chinese characters, *fanqie* distinguishes the pronunciation of one character by introducing two characters.[[2]](#footnote-2) For instance, the pronunciation for east, dōng (东), could be formulized through the combination of two characters, dé (得) and gōng (公), where the first character dé (得) represents the initial sound (assonance) and the second character gōng (公) represents the final (vowel). [[3]](#footnote-3)

Over 500 years later in the twelfth century, the emergence of the *Yunjin* 韵镜 (Late Middle Chinese) is the second significant milestone that Pulleyblank identifies in Chinese historical phonology. The Yunjin is described as the “earliest extant complete rhyme table,” in which it classified 36 distinct initial consonants and listed a corresponding character for each.[[4]](#footnote-4)



**Figure 1**: The 36 Initials of the Yunjing.[[5]](#footnote-5)

Secondly, like the *Qieyun*, the *Yunjing* rhyme table also denoted a clear organization for finals. Initially divided into sixteen shè 摄 (rhyme groups) and later into shān, finals were also delineated in a manner similar to the *Qieyun*, differing by size of groupings and total number of tables. In fact, Pulleyblank notes that “separate but equivalent tables were sometimes necessary in cases where *Qieyun* rhymes had merged in the later language.”[[6]](#footnote-6)

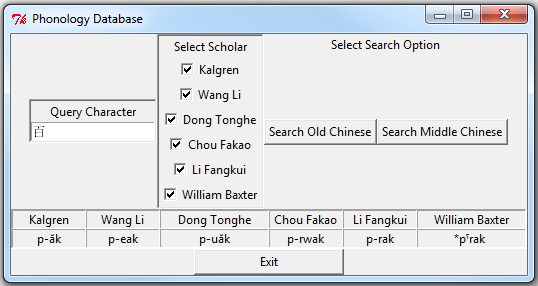
Lastly, the discovery and naming of the four tones stands as a significant milestone for Chinese phonology. These four distinct vowel tones, *píng* 平 (level), *shǎng*上(rising), *qù*去 (departing), and *rù*入 (entering), are what gives Chinese its *tonal* nature and also presents significant challenges for native non-tonal language speakers who wish to learn Chinese. Pulleyblank notes that origin of these tones are somewhat ambiguous. While there lacks any clear indications that the four tones were form Sanskrit origins, its original inception appears to come from lore rather than a well-defined methodology. That is, when a Chinese emperor once asked about the origins of the four tones, the response was simply that each character tone is exemplified through a four-word sentence: *tiānzǐshèngzhé*天子圣哲.[[7]](#footnote-7)

**Research Question**

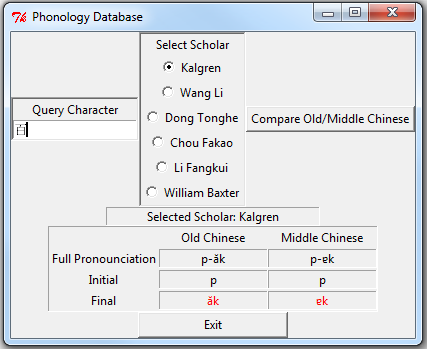
Our database seeks to track such phonological changes over history. Specifically, we seek to identify how various scholars have interpreted the evolution of characters over time. Currently, such records exist in a number of separate rhyme tables and books. By combining these separate records into one single archive, this project serves as a resource that can be used to track the progression of Chinese phonology through specific periods and by specific scholars.

**Prototype Design**

This project focuses on two primary purposes. The first centers on comparing the interpretation of a single character by a number of different scholars. The second centers on comparing a single scholar’s interpretation of a character from Old to Middle Chinese. In Figure 2, an example of the first design is displayed. Given a Chinese character and a set of specific researchers, the application allows users to query for how various scholars interpret the pronunciation of the specified character. In Figure 3, an example of the second design is displayed. Given a particular character and exactly one scholar, the application allows users to query how that particular character changed from Old Chinese to Middle Chinese.



**Figure 2:** A query of how a single character (百) is interpreted by various scholars in a given timeframe (Old or Middle Chinese).



**Figure 3:** A query of how a single character (百) has changed given a single scholar. Note that pronunciations in red-colored font denote how the pronunciation changed between Old and Middle Chinese.

**Discussion & Future Work**

Unfortunately, Python’s Tkinter module cannot be displayed in a web browser or deployed as an executable file. Therefore, application execution is limited to just the command line. Additionally, the current dataset of approximately 100 characters and their associated pronunciations is loaded entirely into memory (a nominal 11KB). This remains one of the central reasons why this “database” was not implemented using traditional database software like MySQL. The overhead of implementing a full-fledged database for simply 11KB proved to be unnecessary. Even with 100 times as many characters (10,000 total characters and associated pronunciations), the total size of the dataset would still only be approximately 1.1MB, which can simply be loaded into memory during application runtime.

Another consideration for future work is scalability; that is, how easy it is for users to add characters. Currently, characters and pronunciations are simply saved in separate text files and are read into the application during runtime. The relations between the characters and their respective pronunciations are inferred by the application (e.g. line 1 of text\_file1.txt corresponds to line 1 of text\_file2.txt). Developing a streamlined and easy-to-understand way for users to add additional characters and pronunciations to the database remains a key improvement for the future.

Lastly, other important expansions include reshaping this project from a standalone Python Tkinter app into a web environment using languages like HTML, PHP, or JavaScript with a true database backend. In order to be easily shared or mounted on a website, this application would have to be implemented using more modern backend software.

**Conclusion**

In sum, this project served as only a first step towards gaining a greater understanding of Chinese historical phonology. Although the current dataset remains small, as more characters are introduced, the relevancy and usefulness of this database will grow proportionally. Using Python’s Tkinter module resulted in a simple GUI that demonstrated one possible solution for front-end design. However, other solutions exist that could potentially increase the visibility of this tool. Moving forward, increasing the size of the current set of characters and shifting towards a web application design stand as the two most relevant next steps.

**Source Code:**

The source code for this project can be found on my Git repo: <https://github.com/bob-zhu/Chinese_Phonology>

**References**

Pulleyblank, Edwin G. “Qieyun and Yunjing: The Essential Foundation for Chinese Historical Linguistics.” *Journal of the American Oriental Society* 118, no. 2 (1998): 200-216.

Pulleyblank, Edwin G. “Chinese Traditional Phonology.” *Asia Major* 12, no. 2 (1999): 101-137.

1. Edwin G. Pulleyblank, “Qieyin and Yunjing: The Essential Foundation for Chinese Historical Linguistics,” *Journal of the American Oriental Society* 118, no. 2 (1998): 202-203. [↑](#footnote-ref-1)
2. Ibid., 202. [↑](#footnote-ref-2)
3. Edwin G. Pulleyblank, “Chinese Traditional Phonology,” *Asia Major* 12, no. 2 (1999): 105. [↑](#footnote-ref-3)
4. Ibid., 115. [↑](#footnote-ref-4)
5. Ibid., 115. [↑](#footnote-ref-5)
6. Ibid., 120. [↑](#footnote-ref-6)
7. Ibid., 108-109. [↑](#footnote-ref-7)