Statistical Analysis for the Combination Ability of Initials and Finals of Chinese Commonly used Characters’ Syllable

Jin Yasheng  
Key Lab of China's National Linguistic Information Technology,Northwest University for Nationalities  
Lanzhou,China  
E-mail: jys2011Y@163.com

Xie Linqing  
Key Lab of China's National Linguistic Information Technology,Northwest University for Nationalities  
Lanzhou,China  
E-mail: xieling8826@sohu.com

Abstract—This paper from the perspective of phonology, using the fundamental theories of phonemics, phonetics and statistics to describe the phonological system of commonly used Chinese characters. Namely to concrete research the combination ability of initials and finals in the phonological system of 2500 commonly used Chinese characters, including 3 aspects statistics and analysis of the frequency distribution of initials and finals, the combination frequency of different initial types and finals of 4 kinds of pronunciation mouth classification, and the combination frequency of different initial types with different final structure types. Through the research some resulted have gained that: 1) the frequency distribution of initials and finals in 2500 commonly used Chinese characters is highly concentrated; 2) the collocation ability of the pursed-mouths in finals of 4 kinds of pronunciation mouth classification with all kinds of initials is weak; 3) the combination ability of zero initial, palatal, affricate, simple vowels and nasal finals is strongest. This paper provides data and certain reference value for the further study on phonological system of commonly used Chinese characters.

Keywords—commonly used Chinese; combination ability; initials; finals; frequency

I. INTRODUCTION

From the perspective of phonology to describe and the research language is a kind of practical and effective method of language study. Prague school advocates the research of language should be focused on the phonemics, namely starting from the analysis of language phoneme, to studies the function of language. Investigate its reason is, on the one hand, compared with other levels of language, the voice level of language is less affected by the relationship between form and content, on the other hand, the phonological system is a relatively closed system, so the study of language phoneme system can dissect the evolution of language system specifically[1]. In the past, the research of phoneme theory often tends to focus on the definition of phonological classified division in different languages and dialects, the acoustic investigation based on the phonological system, as well as the teaching analysis of pronunciation distinguishes from contradiction phonemes between pronunciation distinguishes the righteous teaching analysis and so on. But the research from the aspects of phonological theory to describe and compare the phonemic system of commonly used Chinese characters in the closed space is relatively insufficient, and the research of the statistics and analysis of different categories phoneme combination ability by using the method of statistics and phonology is even more rare. This paper based on the theory of phonology, phonetics and statistics, from 3 aspects of the frequency distribution of initials and finals, the relationship between initials types and finals of 4 kinds of pronunciation mouth classification, the relationship between initial types and final structure. To explore the collocation way and combination ability of phonological system of commonly used Chinese characters, thus provide data and theoretical basis for the further study of whole Chinese phoneme system.

II. RESEARCH METHODS

This paper adopts the theory of structural linguistics, to analysis language from hierarchical structure types, to study the phonological structure of commonly used Chinese, namely to do decomposition research with phonological system of 2500 commonly used Chinese character, to metathesis of its syllables, then using the statistical method for statistical analysis.

A. Overall Processing

This paper bases on the state language work committee in January of 1988 Chinese characters develop in "modern commonly used Chinese character table" (2500) as a sound library, then separate the initial/final and tone of every character's syllable then convert them to corresponding international phonetic alphabet, then using Matlab and SPSS software to statistical analysis the frequency distribution of initials and finals, the combination frequency distribution of initial types and finals of 4 kinds of pronunciation mouth classification, and the combination frequency distribution of initial types and final structure in phonological system of 2500 commonly used Chinese character. Then analysis and study the combination ability and collection way of initials and finals in Chinese phonological system.

B. The Transform of Syllables

Chinese pinyin is only the phonetic notation, but itself does not represent the actual phoneme, need to convert them to international phonetic alphabet the phonemes then they can be used in concrete research. Namely according to the different collocation relationship of initials and finals to establish rules, in order to realize the computer identification and transformation. Conversion consists of three steps: 1) according to the scheme of pinyin, to separate initial/final and tone of Chinese syllable; 2) the reduction of initial and
Namely to restore the situation of shorthand in Chinese phonetic system. Including zero initial's rewrite, the shorthand of finals, and the ellipsis when the collocation of initials and finals due to special collocation rules, all of them were written back into the original; 3) according to their respective corresponding relationship to transform them into international phonetic system [2].

### TABLE I. INTERNATIONAL PHONETIC ALPHABET SYLLABLES CONVERSION TABLES

<table>
<thead>
<tr>
<th>NO.</th>
<th>Characters</th>
<th>Pinyin</th>
<th>Tone</th>
<th>Initial</th>
<th>Final</th>
<th>Syllable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>八</td>
<td>bā</td>
<td>55</td>
<td>p</td>
<td>A</td>
<td>pA</td>
</tr>
<tr>
<td>2</td>
<td>做</td>
<td>zuò</td>
<td>35</td>
<td>ts</td>
<td>uo</td>
<td>tsuo</td>
</tr>
<tr>
<td>3</td>
<td>据</td>
<td>jù</td>
<td>51</td>
<td>tɕ</td>
<td>y</td>
<td>tɕy</td>
</tr>
<tr>
<td>4</td>
<td>常</td>
<td>cháng</td>
<td>35</td>
<td>tʂ</td>
<td>h</td>
<td>tʂh</td>
</tr>
<tr>
<td>5</td>
<td>風</td>
<td>fēng</td>
<td>51</td>
<td>f</td>
<td>əŋ</td>
<td>fəŋ</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>2500</td>
<td>座</td>
<td>zuò</td>
<td>35</td>
<td>ts</td>
<td>uo</td>
<td>tsuo</td>
</tr>
</tbody>
</table>

### III. THE ANALYSIS AND COMPARISON OF FREQUENCY DISTRIBUTION OF INITIALS AND FINALS

There is 22 initials and 38 finals in 2500 commonly used Chinese characters phonemic system, it has the same number of initials and finals in mandarin Chinese, according to analyzed the frequency distribution of initials and finals in 2500 commonly used Chinese word and the compare the quantitative frequency of initials and finals both in 2500 commonly used Chinese characters and standard Chinese, so as to study the characteristics of phonology system of 2500 commonly used Chinese characters.

#### A. Analysis on the Distribution Patterns of Initials and Finals

According to separate the initial, final and tone of every syllable of 2500 commonly used Chinese characters system and statistics the frequency of initials and finals, then gain the frequency distribution pattern of initials and finals in commonly used Chinese characters, as shown in figure 1 and figure 2. Figure 1 and figure 2, the abscissa both represents the frequency of initials and finals, its ordinate both represents the number of initials and finals in different frequency range.

From as shown in figure 1, the initial frequency distribution is very concentrated, the frequency of 15 initials concentrated in the 50-150 range. The highest frequency is zero initial [ŋ], then the next is [tɕ], both of them are only one each, accounted for 13.6% and 8.4% respectively. It is thus clear that the combination ability of the zero initial [ŋ] is significantly higher than other initials. From as shown in figure 2, the frequency distribution of finals is more concentrated, in which 33 finals frequency on a scale of 0-100 range, and the number of finals is the most which distribute in the frequency distribution range between 75-100, is 10, the second is 9 finals which concentrated in frequency distribution range 25 to 50, the highest frequency was simple vowels [i] and [u], which concentrated in the distribution range between 175-225. Overall, the frequency of zero initial is highest in the total initials, because it won’t be restricted by the place of articulation and manner of articulation and can combine with almost all finals, so the frequency of zero initial in the syllables must be high; the high vowels [i], [u] have higher frequency in finals, It is thus clear that the collocation ability of high vowels in simple vowels is stronger.
B. The Comparison of Initial and Final in 2500 Commonly Used Chinese Character with Initial and Final in Standard Chinese

According to the frequency of initial and final in 2500 commonly used Chinese character, and the frequency of initial and final in standard Chinese, within the unit scope of the quantitative comparison, obtained quantitative frequency contrast trend diagram of initial and final in 2500 commonly used Chinese character and initial and final in standard Chinese, as shown in figure 3 and figure 4. In the figure, the abscissa both represents the number of initials and finals, the ordinate represents the quantitative frequency values of them respectively.

From initial frequency quantitative contrast trend diagram, the frequency of initials in commonly used Chinese character compared with the frequency of initials in standard Chinese has great changes, but the highest quantitative frequency is zero initial [¥], the quantitative frequency of only initials [ʂ], [tʂ], [p], [ʨh], [ɕ], [tɕ] and [f] are higher than the quantitative frequency of initials in standard Chinese, then the rest of initials are less than ordinary initials.

From the final frequency quantitative contrast trend diagram, the change of finals in commonly used Chinese character compared with standard Chinese finals quantitative frequency change is more outstanding, and the quantitative frequency values of part of the final in commonly used Chinese character compared with ordinary finals is obvious outstanding embodied in the figure 3, [u], [i], [y], [ɿ] and [ɿ] five peaks. Thus it can be seen, the combination ability of simple vowels is stronger in the finals of commonly used Chinese characters.

In a word, the quantitative frequency of initials and finals in commonly used Chinese characters relative to the quantitative frequency of initials and finals in standard Chinese has a certain differences, no matter the initials in commonly used Chinese characters, or initials in standard Chinese, the of quantitative frequency of zero initial is the highest, the combination ability is strong; The quantitative frequency value of [u], [i], [y], [ɿ] and [ɿ] in simple vowels is obvious and have strong combination ability.

IV. THE RELATIONSHIP BETWEEN INITIAL TYPES AND 4 KINDS FINAL OF PRONUNCIATION MOUTH CLASSIFICATION

According to the place of articulation and manner of articulation, initials can be divided into different types, final is according to the start pronunciation mouth shape are classified into open-months, even-teeth, close-months, and pursed –mouths. By finding out the collocation relations of different initials and 4 kinds final of pronunciation mouth classification, in addition the frequency distribution of them, so as to explore 2500 commonly used Chinese phonology system’s combination ability.

A. The Distributions of Collocation Initial Types and 4 Kinds Final of Pronunciation Mouth Classification

In 2500 commonly used Chinese character, based on the frequency statistics and classification summary of the match of different initial types and 4 kinds final of pronunciation mouth classification, then get the frequency distribution of them, as shown in figure 6 and figure 7. In figure 6 and figure 7, the abscissa represents the initials’ place of articulation and manner of articulation respectively, and the vertical represents the frequency of different places of articulation and manner of articulation initials match the finals in 4 kinds final of pronunciation mouth classification respectively.
As can be seen from the figures above: In the place of articulation of initials, the combination frequency of palatal with even-teeth is the highest, and palatal don’t collocate with open-mouths and close-mouths, it is thus clear that, the combination ability of palatal with even-teeth is the strongest; The types of 4 kinds final of pronunciation mouth classification which blade-palatal, velar, blade-alveolar and labiodental collocate with have similarity on number of frequency distribution to a certain extent. At the same time, they don’t match with even-teeth and pursed-mouths.

In the manner of articulation of initials, the number open-months which plosive, fricative and affricate match with is more. Well even-teeth and pursed-months, open-months and close-months have almost same frequency distribution trend when they match with the initials of different manner of articulation. As the number of nasal and lateral is less, so they have low frequency when they match with 4 kinds final of pronunciation mouth classification compared with other initials.

All in all, in the collocation of 4 kinds final of pronunciation mouth classification compared with other initials.

As can be seen from table 2 and table 3, In the place of articulation of initials, the combination frequency of palatal with even-teeth is the highest, and palatal don’t collocate with open-months and close-months, it is thus clear that, the combination ability of palatal with even-teeth is the strongest; The types of 4 kinds final of pronunciation mouth classification which blade-palatal, velar, blade-alveolar and labiodental collocate with have similarity on number of frequency distribution to a certain extent. At the same time, they don’t match with even-teeth and pursed-mouths.

In the manner of articulation of initials, the number open-months which plosive, fricative and affricate match with is more. Well even-teeth and pursed-months, open-months and close-months have almost same frequency distribution trend when they match with the initials of different manner of articulation. As the number of nasal and lateral is less, so they have low frequency when they match with 4 kinds final of pronunciation mouth classification compared with other initials.

As can be seen from table 2 and table 3, In the place of articulation of initials, the combination frequency of palatal with even-teeth is the highest, and palatal don’t collocate with open-months and close-mouths, it is thus clear that, the combination ability of palatal with even-teeth is the strongest; The types of 4 kinds final of pronunciation mouth classification which blade-palatal, velar, blade-alveolar and labiodental collocate with have similarity on number of frequency distribution to a certain extent. At the same time, they don’t match with even-teeth and pursed-mouths.

In the manner of articulation of initials, the number open-months which plosive, fricative and affricate match with is more. Well even-teeth and pursed-months, open-months and close-months have almost same frequency distribution trend when they match with the initials of different manner of articulation. As the number of nasal and lateral is less, so they have low frequency when they match with 4 kinds final of pronunciation mouth classification compared with other initials.

All in all, in the collocation of 4 kinds final of pronunciation mouth classification compared with other initials.

As can be seen from table 2 and table 3, In the place of articulation of initials, the combination frequency of palatal with even-teeth is the highest, and palatal don’t collocate with open-months and close-mouths, it is thus clear that, the combination ability of palatal with even-teeth is the strongest; The types of 4 kinds final of pronunciation mouth classification which blade-palatal, velar, blade-alveolar and labiodental collocate with have similarity on number of frequency distribution to a certain extent. At the same time, they don’t match with even-teeth and pursed-mouths.

In the manner of articulation of initials, the number open-months which plosive, fricative and affricate match with is more. Well even-teeth and pursed-months, open-months and close-months have almost same frequency distribution trend when they match with the initials of different manner of articulation. As the number of nasal and lateral is less, so they have low frequency when they match with 4 kinds final of pronunciation mouth classification compared with other initials.

All in all, in the collocation of 4 kinds final of pronunciation mouth classification compared with other initials.

As can be seen from table 2 and table 3, In the place of articulation of initials, the combination frequency of palatal with even-teeth is the highest, and palatal don’t collocate with open-months and close-mouths, it is thus clear that, the combination ability of palatal with even-teeth is the strongest; The types of 4 kinds final of pronunciation mouth classification which blade-palatal, velar, blade-alveolar and labiodental collocate with have similarity on number of frequency distribution to a certain extent. At the same time, they don’t match with even-teeth and pursed-mouths.

In the manner of articulation of initials, the number open-months which plosive, fricative and affricate match with is more. Well even-teeth and pursed-months, open-months and close-months have almost same frequency distribution trend when they match with the initials of different manner of articulation. As the number of nasal and lateral is less, so they have low frequency when they match with 4 kinds final of pronunciation mouth classification compared with other initials.

All in all, in the collocation of 4 kinds final of pronunciation mouth classification compared with other initials.

As can be seen from table 2 and table 3, In the place of articulation of initials, the combination frequency of palatal with even-teeth is the highest, and palatal don’t collocate with open-months and close-mouths, it is thus clear that, the combination ability of palatal with even-teeth is the strongest; The types of 4 kinds final of pronunciation mouth classification which blade-palatal, velar, blade-alveolar and labiodental collocate with have similarity on number of frequency distribution to a certain extent. At the same time, they don’t match with even-teeth and pursed-mouths.

In the manner of articulation of initials, the number open-months which plosive, fricative and affricate match with is more. Well even-teeth and pursed-months, open-months and close-months have almost same frequency distribution trend when they match with the initials of different manner of articulation. As the number of nasal and lateral is less, so they have low frequency when they match with 4 kinds final of pronunciation mouth classification compared with other initials.

All in all, in the collocation of 4 kinds final of pronunciation mouth classification compared with other initials.

As can be seen from table 2 and table 3, In the place of articulation of initials, the combination frequency of palatal with even-teeth is the highest, and palatal don’t collocate with open-months and close-mouths, it is thus clear that, the combination ability of palatal with even-teeth is the strongest; The types of 4 kinds final of pronunciation mouth classification which blade-palatal, velar, blade-alveolar and labiodental collocate with have similarity on number of frequency distribution to a certain extent. At the same time, they don’t match with even-teeth and pursed-mouths.

In the manner of articulation of initials, the number open-months which plosive, fricative and affricate match with is more. Well even-teeth and pursed-mouths, open-months and close-mouths have almost same frequency distribution trend when they match with the initials of different manner of articulation. As the number of nasal and lateral is less, so they have low frequency when they match with 4 kinds final of pronunciation mouth classification compared with other initials.

All in all, in the collocation of 4 kinds final of pronunciation mouth classification compared with other initials.

As can be seen from table 2 and table 3, In the place of articulation of initials, the combination frequency of palatal with even-teeth is the highest, and palatal don’t collocate with open-months and close-mouths, it is thus clear that, the combination ability of palatal with even-teeth is the strongest; The types of 4 kinds final of pronunciation mouth classification which blade-palatal, velar, blade-alveolar and labiodental collocate with have similarity on number of frequency distribution to a certain extent. At the same time, they don’t match with even-teeth and pursed-mouths.

In the manner of articulation of initials, the number open-months which plosive, fricative and affricate match with is more. Well even-teeth and pursed-mouths, open-months and close-mouths have almost same frequency distribution trend when they match with the initials of different manner of articulation. As the number of nasal and lateral is less, so they have low frequency when they match with 4 kinds final of pronunciation mouth classification compared with other initials.

All in all, in the collocation of 4 kinds final of pronunciation mouth classification compared with other initials.

As can be seen from table 2 and table 3, In the place of articulation of initials, the combination frequency of palatal with even-teeth is the highest, and palatal don’t collocate with open-months and close-mouths, it is thus clear that, the combination ability of palatal with even-teeth is the strongest; The types of 4 kinds final of pronunciation mouth classification which blade-palatal, velar, blade-alveolar and labiodental collocate with have similarity on number of frequency distribution to a certain extent. At the same time, they don’t match with even-teeth and pursed-mouths.

In the manner of articulation of initials, the number open-months which plosive, fricative and affricate match with is more. Well even-teeth and pursed-mouths, open-months and close-mouths have almost same frequency distribution trend when they match with the initials of different manner of articulation. As the number of nasal and lateral is less, so they have low frequency when they match with 4 kinds final of pronunciation mouth classification compared with other initials.

All in all, in the collocation of 4 kinds final of pronunciation mouth classification compared with other initials.

As can be seen from table 2 and table 3, In the place of articulation of initials, the combination frequency of palatal with even-teeth is the highest, and palatal don’t collocate with open-months and close-mouths, it is thus clear that, the combination ability of palatal with even-teeth is the strongest; The types of 4 kinds final of pronunciation mouth classification which blade-palatal, velar, blade-alveolar and labiodental collocate with have similarity on number of frequency distribution to a certain extent. At the same time, they don’t match with even-teeth and pursed-mouths.

In the manner of articulation of initials, the number open-months which plosive, fricative and affricate match with is more. Well even-teeth and pursed-mouths, open-months and close-mouths have almost same frequency distribution trend when they match with the initials of different manner of articulation. As the number of nasal and lateral is less, so they have low frequency when they match with 4 kinds final of pronunciation mouth classification compared with other initials.

All in all, in the collocation of 4 kinds final of pronunciation mouth classification compared with other initials.
VI. CONCLUDING REMARKS

This paper is based on the statistical study of 2500 commonly used Chinese characters, from the perspective of phonetics, phonology and statistics, to concrete analysis and statistics the frequency of initials and finals, the combination frequency of different types of initials and finals. According to the study of the frequency distribution of initials and finals, the collocation situation of initial types with 4 kinds final of pronunciation mouth classification, and initial types with final structures. We can gain that the frequency of initials and finals in 2500 commonly used Chinese characters is concentrated on a whole. The combination ability of pursed-mouths of 4 kinds final of pronunciation mouth classification, with all kinds of initials is weak; The combination ability of zero initial, palatal, affricate, simple vowels and nasal finals are strongest. The research of this paper basically reflects the structure distribution patterns of the combination of initials and finals in commonly used Chinese character system, also laid the foundation for the further study of Chinese phonology structure, providing the certain reference value.

REFERENCES