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

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

Xie Linqing<br>Key Lab of China's National Linguistic Information<br>Technology,Northwest University for Nationalities Lanzhou, China<br>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 $\mathbf{2 5 0 0}$ commonly used Chinese characters is highly concentrated;2)the collocation ability of the pursedmouths 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 specificly[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 4 kinds final 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
final；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

| NO． | characters | pinyin | tone | IPA |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | initial | final | syllable |
| 1 | 八 | bā | 55 | p | A | pA |
| 2 | 做 | zuò | 35 | ts | uo | tsuo |
| 3 | 聚 | jù | 51 | t6 | y | t6y |
| 4 | 常 | cháng | 35 | tsh | an | tşhay |
| 5 | 凤 | fèng | 51 | f | ә］ | fəŋ |
| $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | ．． | $\ldots$ | $\ldots$ |
| 2500 | 座 | zuò | 35 | ts | uo | tsuo |

## 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．


Figure 1．The frequency distribution pattern of initials


Figure 2．The frequency distribution pattern of finals
From as shown in figure1，the initial frequency distribution is very concentrated，the frequency of 15 initials concentrated in the $50-150$ range，The highest frequency is zero initial［ $¥$ ］，then the next is［tc ］，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 $[¥]$ 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 combinate 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.


Figure 3. the initials quantitative frequency comparison


Figure 4. the finals quantitative frequency comparison
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 grate changes ,but the highest quantitative frequency is zero initial $[¥]$, the quantitative frequency of only initials $[¥ \ddot{\square}],[\mathrm{s}],[\mathrm{ts}],[\mathrm{p}],[\mathrm{tch}],[6],[\mathrm{t} 6]$ 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 $3,[\mathrm{u}],[\mathrm{i}],[\mathrm{y}],[\mathrm{l}]$ and $[\mathrm{l}]$ 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 $[\mathrm{u}],[\mathrm{i}],[\mathrm{y}],[\mathrm{l}]$ and $[\mathrm{n}]$ 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-mouths, even-teeth, close-mouths, 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 differents place of articulation and manner of articulation initials match the finals in 4 kinds final of pronunciation mouth classification respectively.


Figure 5. The collocation frequency distribution of initial's place of articulation and 4 kinds final of pronunciation mouth classification


Figure 6. the collocation frequency distribution of initial's manner of articulation and 4 kinds final of pronunciation mouth classification

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 openmouths which plosive, fricative and affricate match with is more. Well even-teeth and pursed-mouths, open-mouths 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 with different initial types ,the collocation ability of pursed-mouths is the weakest compared with other three. This is because the pursedmouths itself only have [y], [y\&], [yan], [yn] and [yn] five, and due to various constraints of collocation rules of initial and final, only a [n], [1], [t6], [tch], [6] and zero initial [ $¥$ ] can match with them, so combination ability of them is weak.

## V. THE RELATIONSHIP BETWEEN INITIAL TYPES AND FINALS STRUCTURE

Finals can be divided into simple vowels, diphthong and nasal finals with their respective structure.In 2500 commonly used Chinese characters, according to statistic and classification summary,concluded the specific frequency when different initial types match the finals structure, and calculate its weight distribution, as shown in table II and table III.

TABLE II. THE WEIGHT DISTRIBUTION OF COMBINATION OF ARTICULATION PALCE OF INITIALS WITH FINAL STRUCTURES (\%)

| Final <br> structure <br> Articulation <br> place | simple <br> vowels | diphthong | nasal finals |
| :--- | :---: | :---: | :---: |
| bilabial | 4.4 | 3.3 | 4.1 |
| labiodental | 1.4 | 0.4 | 1.8 |
| blade-alveolar | 2.3 | 2.6 | 2.3 |
| supradental | 4.8 | 5.3 | 7.2 |
| blade-palatal | 7.2 | 3.6 | 7 |
| palatal | 4.8 | 5.5 | 8.2 |
| velar | 2.7 | 4.4 | 4.6 |
| 0 initial | 3.5 | 3.9 | 4.8 |

TABLE III. THE WEIGHT DISTRIBUTION OF COMBINATION OF articulation manner of initials with final structures (\%)

| Final <br> stcucture <br> Articulation <br> manner | simple <br> vowels | diphthong | nasal finals |
| :---: | :---: | :---: | :---: |
| plosive | 6.9 | 7.5 | 9.8 |
| fricative | 7.8 | 6.5 | 9.5 |
| affricate | 8.9 | 7.6 | 11.4 |
| nasal | 2 | 1.9 | 1.9 |
| lateral | 2 | 1.6 | 2.5 |
| 0 initial | 3.5 | 3.9 | 4.8 |

As can be seen from table 2 and table 3, In the place of articulation the highest weight is the combination of palatal with nasal finals accounted for $8.2 \%$, the combination ability is the strongest; The next is the combination of supradental with nasal finals and blade_palatal with simple vowels ,their weight both account for $7.2 \%$, on the whole, the combination ability of labiodental with different final structures is weaker, the combination ability of labiodental with diphthong is the weakest.

In the manner of articulation, the combination weight of affricate with all the final structures is the highest,so the combination ability of is the strongest; The combination weight of Nasal and lateral with all final structures is average. The combination weight of plosive, fricative, affricate with all the final structures is much higher than nasal, lateral and zero initial.

Overall, the weight of nasal finals is higher than simple vowels and diphthong, embodied in table 3.This is because the nasal finals itself has 16 , has a large proportion in the finals, and evenly distributed in 4 kinds final of pronunciation mouth classification,the initials of different manner of articulation can match with 4 kinds final of pronunciation mouth classification, nasal finals has the strongest combination ability.

## 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 pursedmouths 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

[1]. Dinoj Surendran,Analysis and Automatic Recognition of Tones in Mandarin Chinese,2007. pp.100-110.
[2]. Pass P.Outlines of comparative phonetics [M].Shanghai: The Commercial Press,1930,pp30-40.
[3]. Lieberman P.Intonation,perception, and language[M].Cambrige:The MIT Press,1967. pp.100-110
[4]. Dinoj Surendran,Analysis and Automatic Recognition of Tones in Mandarin Chinese,2007. pp.101-120.
[5]. Greenberg,H.H. A method of measuring functional yield as applied to tone in Mrican languages. Georgetown University Monograph Series on Languages and Linguistics Dec.1959. pp.7-16
[6]. Hockett,C.F. A Manual of Phonology. Baltimore Waverly Press,1955. pp.218-220.
[7]. Liebeman P.The biology and evolution of language[M]. Cambrige:Harvard University Press, 1984
[8]. Ron J.Baken et al.Chest wall movements prior to phonation[J]. Journalof Speech and Hearing Research. pp. 862-872
[9]. Ladefoged $P$. Three areas of experimental phonetics[M]. London:Oxford University Press.1967. pp. 862-872

