A Computational Method for Phonemic Functional Load Based on Linguistic Structural Function

Authors: Kong Jiangping, Li Yonghong, Kong Jiangping

Date: 2019-06-20T00:00:00+00:00

Abstract

Based on the phonemic opposition theory of structural linguistics and grounded in various language corpora, this study conducts a series of theoretical investigations and data analyses to propose a computational methodology for the functional load of phonological structure. The computational procedure primarily comprises: 1) statistical analysis of syllabic opposition frequency; 2) calculation of the functional load of opposition types; 3) calculation of the functional load of the phonemic system; and 4) calculation of the average functional load of phonemes. Employing this methodology, with 3,000 syllables from 20 Chinese dialects as the research corpus, we calculate the distribution of opposition type load, phonemic system load, and average phoneme load across dialects, and establish the foundational theory for phonemic load computation. This methodology is applicable to research on phonemic function, historical sound change and evolution, as well as applied research in speech engineering.

Full Text

Preamble

A Calculation Method for Phoneme Load Based on Linguistic Structural Function*

Kong Jiangping¹ and Li Yonghong²

2. Institute of Chinese Ethnic Information Technology, Northwest Minzu University

Abstract

Based on phoneme opposition theory in structural linguistics and grounded in various language corpora, this study conducts a series of theoretical investigations and data analyses to propose a computational method for phoneme structural-functional load. The calculation procedure primarily includes: 1) statistical analysis of syllable opposition frequencies; 2) calculation of load quantities for opposition types; 3) calculation of load for the phonemic system; and 4) calculation of average phoneme load. Applying this method to 3,000 syllables across 20 Chinese dialects, we compute the distribution of opposition type load, phonemic system load, and average phoneme load across dialects, and propose a foundational theory for phoneme load calculation. This method can be applied to research on phoneme function, historical sound change and language evolution, as well as speech engineering applications.

Keywords: phoneme, functional load, phoneme load, phoneme structure

Introduction

The phoneme is a core concept in structuralist phonemics, referring to the smallest phonological unit that distinguishes meaning in a language. Establishing a phonemic system for a language relies primarily on principles of opposition, complementation, similarity, and economy (Bloomfield 1980; Sapir 1985; Bloch & Trager 1965). Structuralist phonemics has played a crucial role in basic linguistic description and historical research, forming the foundation of modern linguistics. However, due to inherent differences in phonemic systems across languages, structuralist phonemics exhibits clear limitations in explaining phoneme function. For instance, Sino-Tibetan languages are predominantly monosyllabic, composed of initials, finals, and tones, whereas Indo-European languages are primarily polysyllabic, constructed from phonemes. Moreover, some phonemes in a language may appear in only a handful of words with poor systematicity. These observations reveal that the structuralist definition of phoneme inadequately explains the frequency and functional role of phonemes within a language.

This directly leads to difficulties in using current phonemic theory to explain numerous issues in language information transmission, language evolution, and linguistic typology. Additionally, in phonemic description and classification, some scholars adopt a physiological articulatory perspective, others emphasize acoustic properties using distinctive acoustic features, while still others stress psychological functions based on perceptual results. Consequently, various divergent viewpoints and debates exist in phonemic research. To reconcile these disputes, Mr. Chao Yuen-ren specifically discussed "The Non-Uniqueness of Phonemic Solutions" (Chao 1934/1985).

These theoretical issues in phonemics have long concerned linguists, among which the study of phoneme functional load represents an attempt to address these problems. The concept and research of functional load can be traced back to the early Prague School period (Mathesius 1929; Jakobson 1931; Trubetzkoy 1939), which primarily focused on binary phonological oppositions. In the 1950s, major contributions included studies by Hockett (1955, 1967) and Greenberg (1959). Hockett argued that the importance of functional load lies in its value for describing phonological systems, providing a metric for understanding language information, redundancy, and speech recognition. Greenberg contended that functional load reflects, in a general way, the contribution of individual members of a set of phonemes or distinctive features to signaling meaningful distinctions.

In the 1960s, Hoenigswald (1960) conducted research on functional load and sound change, proposing the hypothesis that "in a language, if an opposition is rarely used, its disappearance poses less danger to the system than the loss of an opposition with high functional load." Also in the 1960s, Professor Wang Shiyuan (Wang 1967) conducted seminal research on functional load, as did King (1965, 1967a, 1967b). Professor Wang Shiyuan first implemented computational methods for functional load, identified difficulties in quantifying it, and provided solutions. He discussed three common distributions in phonemic systems, measurement methods by Hockett and Greenberg, and their relationships with communication theory by Shannon and Weaver (1949; Shannon 1951) and Kucera (1963), as well as various linguistic concepts, systematically developing four methods for quantifying functional load. Professor Wang' s research established a fundamental theoretical framework for subsequent functional load studies. He also noted that "if functional load does indeed play a role in sound change, then quantitative explanation could at least partially illuminate this difficult problem." Robert King studied sound change in conjunction with functional load, focusing particularly on the relationship between phoneme function and phonetic change.

At the turn of the 21st century, research by Surendran and Niyogi (2003) and Surendran and Levow (2004) advanced the field. In their study, Surendran and Levow discussed Hockett' s definition and examined the functional load of phonemes, distinctive features, and suprasegmental features. They also investigated the functional load of Chinese tones, finding that it is as high as that of vowels.

Research on phoneme functional load has promoted the study of linguistic function and advanced computational linguistics and speech technology. For example, diphones and triphones commonly used in current speech recognition and synthesis originate from functional load theory. Previous studies have conducted statistical and quantitative analyses of phoneme functional load based on large texts, calculating language entropy and redundancy. However, most languages in the world lack writing systems and textual documentation, which severely limits the application and development of this theoretical approach in linguistic research. Therefore, developing a method that can calculate and quantify phoneme function using only basic morphemes and vocabulary has become urgently necessary.

Based on phoneme opposition theory in structural linguistics and corpora of Mandarin Chinese, Chinese dialects, Tibetan dialects, and Tibeto-Burman languages, we have conducted a series of theoretical studies and data analyses, ultimately proposing a computational method for phoneme structural-functional load. This "structural-functional load" within phonemes is closed in nature. Structurally, it better reflects the properties of a language' s phonemic system. Since only basic morphemes are required for research, it can be applied to phoneme function studies, historical sound change and evolution research, and speech synthesis and recognition studies for any language, with the hope of establishing a new theoretical framework for phoneme function research. This paper uses Chinese dialects as examples to introduce the specific calculation methods for phoneme structural-functional load.

I. Definition and Types of Phoneme Load

1.1 Definition

In defining phoneme load, we define the distinction between two monosyllabic morphemes as one load unit of language, borne by one or more different phonemes. While most languages are described using phonemic phonemes, Sino-Tibetan languages in China have an additional phonemic level of initials, finals, and tones in their phonological structure. Phonemic units can be either phonemic phonemes or initial/final/tone phonemes. To better explain the calculation method for phoneme load and the phoneme load of Sino-Tibetan languages, we select initials, finals, and tones as the units for phoneme load calculation. If one phonemic unit distinguishes two morphemes, that unit bears the phoneme load of those morphemes; if multiple phonemic units distinguish them, the phoneme load is shared among these units.

The load within a phonemic system concerning phoneme structure, distribution, and function is termed "phoneme structural-functional load" (音位结构功能负担), abbreviated as "phoneme load" (音位负担). Research on phoneme structural-functional load primarily investigates the internal phoneme structure, distribution, and functional load of basic morphemes in a language. "Speech functional load research" (言语功能负担研究), which investigates phoneme functional load based on actual language texts, is fundamentally distinct from our approach.

To clearly explain the calculation method, we propose several new concepts based on the monosyllabic characteristics of Sino-Tibetan languages and provide detailed explanations.

Morpheme: In this paper, refers to monosyllabic morphemes.

Syllable opposition: Refers to the opposition between different monosyllabic morphemes, manifested as initial opposition, final opposition, tone opposition, or combinations thereof between two monosyllabic morphemes.

Language load unit: One monosyllabic morpheme bears one language load unit.

Syllable load: Within a closed phonemic system and basic morpheme system, the load of a syllable equals the sum of its load, i.e., the number of homophones for that syllable.

1.2 Types

In Sino-Tibetan languages, basic morphemes are mostly monosyllabic, with each syllable composed of initials and finals, or initials, finals, and tones. Between syllables, four opposition types and eight opposition methods emerge, including: one three-way opposition, three two-way oppositions, three single oppositions, and one non-opposition, forming the basic form and framework of phonemic system structure and distribution, as shown in Table 1 .

Table 1 Phoneme Opposition Type Table - Three-way opposition (initial/final/tone) - Two-way oppositions (initial/final, initial/tone, final/tone) - Single oppositions (initial, final, tone) - No opposition (initial/final/tone identical)

We use Beijing dialect as an example to explain these eight opposition methods. Three-way opposition occurs when two monosyllabic morphemes differ in initial, final, and tone, e.g., 八 pa55 vs. 笛 ti35. Two-way opposition occurs when two monosyllabic morphemes differ in only two phonemic units: initial/final opposition (e.g., 八 pa55 vs. 低 ti55), initial/tone opposition (e.g., 八 pa55 vs. 大 ta51), or final/tone opposition (e.g., 八 pa55 vs. 不 bu51). Single opposition occurs when two monosyllabic morphemes differ in only one phonemic unit: initial opposition (e.g., 八 pa55 vs. 搭 ta55), final opposition (e.g., 八 pa55 vs. 波 bo55), or tone opposition (e.g., 八 pa55 vs. 拔 pa35). Non-opposition occurs when two monosyllabic morphemes have identical phonemic units, i.e., homophones (e.g., 八 pa55 vs. 巴 pa55).

Based on these definitions, we discuss the calculation methods for phoneme load and the entire linguistic framework in the following section.

II. Calculation Methods

Phoneme structural load calculation is manifested through opposition relationships between syllables, including syllable opposition frequency statistics, opposition type load calculation, and phoneme load calculation. Syllable opposition frequency statistics form the foundation for both opposition type load calculation and phoneme load calculation.

2.1 Syllable Opposition Frequency Calculation

This step calculates the opposition relationship between each morpheme' s corresponding syllable and all other morpheme syllables, i.e., opposition frequency. Below, we use the 3,000-character Beijing dialect dataset from the *Chinese Dialect Character Compendium* as an example to illustrate the calculation of syllable opposition for each morpheme. Figure 1 [Figure 1: see original paper] shows the opposition frequency calculation diagram for the morpheme “巴 (pa55).”

From Figure 1, we can see that “巴” forms 2,999 opposition relationships with other morphemes among the 3,000 morphemes, of which 2,997 form different syllables (oppositions) and 2 form homophones of “巴” (non-oppositions), specifically:

1. **Three-way opposition:** Initial/final/tone opposition occurs 2,120 times, meaning 2,120 morphemes differ from “巴” in initial, final, and tone.
2. **Two-way opposition:** Initial/final opposition occurs 680 times (same tone, different initial and final); initial/tone opposition occurs 42 times (same final, different initial and tone); final/tone opposition occurs 98 times (same initial, different final and tone).
3. **Single opposition:** Initial opposition occurs 17 times (same final and tone, different initial); final opposition occurs 33 times (same initial and tone, different final); tone opposition occurs 7 times (same initial and final, different tone).

After completing the calculation for the first syllable “巴,” the same method is applied to calculate the second syllable, third syllable, and so on, until all 3,000 syllables are processed, yielding a syllable opposition frequency table. Table 2 shows the syllable opposition frequency table for the 3,000-character Beijing dialect dataset.

Table 2 Beijing Dialect Syllable Opposition Frequency Table [Table structure with columns for: Serial No., Character, Syllable, Three-way, Initial/Final, Initial/Tone, Final/Tone, Initial, Final, Tone]

2.2 Opposition Type Load Calculation

Opposition type load calculation includes calculations for three-way, two-way, and single opposition types, with the magnitude indicating the relationships between syllables and the combinatorial structural patterns of phonemes in a language. The calculation process involves: (1) merging homophones by syllable, where each syllable’ s load equals its number of homophones T(i); (2) multiplying each opposition frequency by the T(i) coefficient since homophones share identical opposition frequencies; (3) distributing each syllable’s load across the three-way, two-way, and single opposition conditions based on proportional relationships among opposition frequencies.

Additionally, the total number of monosyllabic morphemes in the closed corpus is denoted by symbol N, representing the total load of the entire corpus.

(1) Three-way Opposition Phoneme Load Calculation

The load quantity for a syllable’s three-way opposition type is calculated through the number of three-way opposition frequencies. Therefore, the three-way opposition load FFL_TR(i) for the i-th syllable can be obtained through Formula (1):

$$\text{FFL\_TR}(i) = \frac{\text{SYD}(i) \times T(i)}{N - T(i)}$$

where SYD represents three-way opposition frequency.

(2) Two-way Opposition Phoneme Load Calculation

The two-way opposition phoneme load for a syllable equals the sum of loads for the three two-way opposition methods. Therefore, the two-way opposition load FFL_DB(i) for the i-th syllable can be obtained through Formula (2):

$$\mathrm{FFL\_DB}(i) = \frac{(\mathrm{SY}(i) + \mathrm{SD}(i) + \mathrm{YD}(i)) \times T(i)}{N - T(i)}$$

where SY represents initial/final two-way opposition frequency, SD represents initial/tone two-way opposition frequency, and YD represents final/tone two-way opposition frequency.

(3) Single Opposition Phoneme Load Calculation

The single opposition type load for a syllable equals the sum of loads for the three single opposition methods. Therefore, the single opposition type load FFL_SG(i) for the i-th syllable can be obtained through Formula (3):

$$\mathrm{FFL\_SG}(i) = \frac{(\mathrm{S}(i) + \mathrm{Y}(i) + \mathrm{D}(i)) \times T(i)}{N - T(i)}$$

where S represents single initial opposition frequency, Y represents single final opposition frequency, and D represents single tone opposition frequency.

Through the above formulas for opposition type load calculation, we can obtain the load quantities for different syllable opposition types and differences among syllables in opposition type distribution. Using the syllable "(pa55)" as an example, which corresponds to three morphemes "巴, 疤, 八," the total load of "(pa55)" is 3, distributed across the seven opposition conditions (excluding "no opposition") according to proportional relationships among opposition frequencies, as shown in Table 3 .

Table 3 Beijing Dialect Opposition Type Load Calculation Table [Table structure with columns for: Serial No., Syllable, Homophones, Load, Three-way, Initial/Final, Initial/Tone, Final/Tone, Initial, Final, Tone]

2.3 Phoneme Load Calculation

In a linguistic system, each syllable' s status is manifested through eight opposition relationships generated by three opposition units: initial, final, and tone. The more homophones a syllable has, the greater its phoneme load and the higher its status in the linguistic system, resulting in greater load distribution to its initial, final, and tone.

(1) Initial Load Calculation

In three-way opposition distribution, the initial bears 1/3 of the total load; in two-way initial/final and initial/tone distributions, the initial bears 1/2 of the total load each; in single initial opposition, the initial bears the entire load. Therefore, initial load FFL_S(i) can be calculated through Formula (4):

$$\text{FFL\_S}(i) = \frac{\text{SYD}(i)/3 + (\text{SY}(i) + \text{SD}(i))/2 + \text{S}(i)}{N - T(i)} \times T(i)$$

Then, loads for identical initials are summed to obtain a particular initial' s load FFL_S, where P represents the number of syllables containing that initial, as shown in Formula (5):

$$\text{FFL\_S} = \sum_{i=1}^{P} \text{FFL\_S}(i)$$

The average load for a particular initial FFL_S_AV is obtained by dividing its load by P, as shown in Formula (6):

$$\text{FFL\_S\_AV} = \frac{\sum_{i=1}^{P} \text{FFL\_S}(i)}{P}$$

Other initials are calculated using the same method to obtain each initial' s load. The sum of all initial loads constitutes the total initial load.

(2) Final Load Calculation

In three-way opposition distribution, the final bears 1/3 of the total load; in two-way initial/final and final/tone distributions, the final bears 1/2 of the total load each; in single final opposition, the final bears the entire load. Therefore, final load FFL_Y(i) can be calculated through Formula (7):

$$\text{FFL\_Y}(i) = \frac{\text{SYD}(i)/3 + (\text{SY}(i) + \text{YD}(i))/2 + \text{Y}(i)}{N - T(i)} \times T(i)$$

Then, loads for identical finals are summed to obtain a particular final' s load FFL_Y, where P represents the number of syllables containing that final, as shown in Formula (8):

$$\text{FFL\_Y} = \sum_{i=1}^{P} \text{FFL\_Y}(i)$$

The average load for a particular final FFL_Y_AV is obtained by dividing its load by P, as shown in Formula (9):

$$\mathrm{FFL\_Y\_AV} = \frac{\sum_{i=1}^{P} \mathrm{FFL\_Y}(i)}{P}$$

Other finals are calculated using the same method to obtain each final' s load. The sum of all final loads constitutes the total final load.

(3) Tone Load Calculation

In three-way opposition distribution, the tone bears 1/3 of the total load; in two-way initial/tone and final/tone distributions, the tone bears 1/2 of the total load each; in single tone opposition, the tone bears the entire load. Therefore, tone load FFL_D(i) can be calculated through Formula (10):

$$\mathrm{FFL\_D}(i) = \frac{\mathrm{SYD}(i)/3 + (\mathrm{SD}(i) + \mathrm{YD}(i))/2 + \mathrm{D}(i)}{N - T(i)} \times T(i)$$

Then, loads for identical tones are summed to obtain a particular tone' s load FFL_D, where P represents the number of syllables containing that tone, as shown in Formula (11):

$$\mathrm{FFL\_D} = \sum_{i=1}^{P} \mathrm{FFL\_D}(i)$$

The average load for a particular tone FFL_D_AV is obtained by dividing its load by P, as shown in Formula (12):

$$\mathrm{FFL\_D\_AV} = \frac{\sum_{i=1}^{P} \mathrm{FFL\_D}(i)}{P}$$

Other tones are calculated using the same method to obtain each tone' s load. The sum of all tone loads constitutes the total tone load.

III. Calculation of Phoneme Structural Load in Chinese Dialects

Based on the phoneme load algorithm described in the previous section, we calculated the phoneme structural load for 20 Chinese dialects. This section uses these results to illustrate the linguistic significance of phoneme load. The procedure involves: (1) statistical calculation of syllable opposition frequencies to obtain each syllable' s opposition frequencies across different opposition types; (2) calculation of opposition type loads to determine the load borne by each opposition method within each opposition type, with loads converted to percentages for convenience (the same applies below), as shown in Figure 2 [Figure 2: see original paper] and Table 3; (3) calculation of phoneme loads (initials, finals, tones), as shown in Figure 3 [Figure 3: see original paper]; (4) calculation of average loads by dividing each dialect' s initial load by its number of initials, with the same method applied to finals and tones, as shown in Figure 4 [Figure 4: see original paper].

3.1 Syllable Opposition Type Frequency in Chinese Dialects

Figure 2 Percentage Relationship Diagram of Different Opposition Methods

From Figure 2, we can observe: (1) Although single opposition is fundamental for meaning distinction, its value is very small; Chinese dialects rely more on three-way and two-way oppositions to differentiate meaning; (2) Opposition method loads have relatively stable ranges; (3) Two-way and three-way oppositions exhibit an inverse relationship.

Table 4 Frequency Percentage of Different Opposition Methods (%)
[Table with columns: Dialect, Three-way, Two-way, Single, Dialect, Three-way, Two-way, Single]

From Table 4 data, we can conclude: (1) Among three-way oppositions, Wenzhou dialect shows the highest value at 78.38%, while Taiyuan shows the lowest at 65.63%, with a distribution range of 12.75 percentage points and an average of 72%; (2) Among two-way oppositions, Taiyuan shows the highest value at 31.04%, while Wenzhou shows the lowest at 19.49%, with a distribution range of 11.55 percentage points and an average of 25.27%; (3) Among single oppositions, Taiyuan shows the highest value at 3.33%, while Guangyun shows the lowest at 1.61%, with a distribution range of only 1.72 percentage points and an average of 2.47%.

In phonological theory, determining a phoneme typically depends on the existence of minimal pairs—the principle of phonological opposition. However, examining Chinese dialect opposition types reveals that the function of minimal opposition is very weak. Therefore, using minimal opposition to describe Chinese dialects fails to capture their phonemic function. The second principle in phonemics is complementation. While two-way and three-way oppositions include complementation phenomena, most are not caused by complementation. In Chinese, the pairs ge/ke/he and ji/qi/xi are not actually handled using the complementation principle either. Utilizing opposition type frequencies can describe not only phoneme opposition types and distribution but also phoneme function types, opening new avenues and research methods for studying phoneme structure, information function, language contact and convergence, and historical evolution.

3.2 Total Phoneme Load in Chinese Dialects

Figure 3 Total Load Relationship of Initials, Finals, and Tones in Chinese Dialects

From Figure 3, we can observe: (1) Initials and finals in Chinese dialects bear relatively large loads, with finals slightly exceeding initials, while tones bear smaller loads; (2) Total loads for initials, finals, and tones each have certain ranges across dialects; (3) Total loads for initials and finals are inversely related to total tone load—greater initial/final loads correspond to smaller tone loads.

Table 5 Total Phoneme Load Percentage (%) [Table with columns: Dialect, Initial Load, Final Load, Tone Load, Dialect, Initial Load, Final Load, Tone Load]

From Table 5, we can see: (1) Among Chinese dialects, Guangyun shows the highest initial load at 37.07%, while Chaozhou shows the lowest at 33.65%, with a distribution range of 3.42 percentage points and an average of 35.35%; (2) For final load, Guangyun shows the highest at 38.37%, while Wenzhou shows the lowest at 34.9%, with a distribution range of 3.47 percentage points and an average of 36.89%; (3) For tone load, Guangzhou shows the highest at 30.01%, while Guangyun shows the lowest at 24.56%, with a distribution range of 5.45 percentage points and an average of 27.76%.

Examining total loads for initials, finals, and tones across Chinese dialects reveals differences. First, total loads for initials, finals, and tones vary, reflecting differences in phoneme structure and function. Second, initials and finals share similar properties; overall, final loads are slightly greater than initial loads, while tone loads are substantially smaller than both. Third, quantitatively, larger initial and final loads correspond to smaller tone loads. These data demonstrate that initials, finals, and tones in Chinese dialects maintain a dynamic and organic relationship, simultaneously reflecting changes in Chinese dialect phoneme structure. Since identical morphemes are selected across dialects, the amount of information to be expressed is the same. However, during long-term evolution, various factors have influenced each dialect' s phoneme structure system. Therefore, phoneme load description provides precise digital characterization of dynamic phonemic systems and functions.

3.3 Average Phoneme Load in Chinese Dialects

Figure 4 Average Phoneme Load Relationship in Phonemic Systems

From Figure 4, we can observe: (1) Average loads for initials and finals in Chinese dialects are relatively small, with initials slightly exceeding finals, while tones show larger average loads; (2) Average loads for initials, finals, and tones each have certain ranges; (3) Average tone load is inversely related to average initial and final loads—greater tone average loads correspond to smaller initial and final average loads.

Table 6 Average Phoneme Load Percentage (%) [Table with columns: Dialect, Initial Avg., Final Avg., Tone Avg., Dialect, Initial Avg., Final Avg., Tone Avg.]

From Table 6, we can see: (1) Among Chinese dialects, Guangzhou shows the highest initial average load at 33.23%, while Guangyun shows the lowest at 12.61%, with a distribution range of 20.62 percentage points and an average of 22.92%; (2) For final average load, Wenzhou shows the highest at 18.88%, while Guangyun shows the lowest at 3.87%, with a distribution range of 15.01 percentage points and an average of 11.38%; (3) For tone average load, Guangyun shows the highest at 83.53%, while Guangzhou shows the lowest at 57.37%, with a distribution range of 21.16 percentage points and an average of 70.45%.

Examining the properties of average loads for initials, finals, and tones in Chinese dialects reveals several patterns. First, tone average loads show relatively large values. Second, initial and final average loads are smaller, with initial average loads slightly exceeding final average loads. Third, tone average loads are inversely related to initial and final average loads. Overall, average load exhibits opposite properties to total load: dialects with high total loads have smaller average loads, while languages with small total loads have larger average loads. For example, Guangzhou dialect has many tones and a relatively large total tone load, but when averaged, Guangzhou' s tone average load is actually smaller than Beijing' s. Detailed analysis reveals that Guangzhou tones and initials/finals exhibit extensive complementary distribution; although both dialects use 3,000 morphemes, the actual tonal function is weaker.

Average load differs from total load in that total load relates to phoneme quantity, whereas average load is unaffected by phoneme quantity. Within a phonemic system, average load represents the burden per phoneme. Larger phoneme average loads indicate more morphemes to be distinguished, which also leads to more homophones. The focus of this paper is to explain the basic definitions and algorithms of phoneme structural load without delving into detailed content regarding Chinese dialect phoneme loads. In reality, within a phonemic system, each phoneme's total load and average load differ, generally following an approximately exponential distribution (Kong 2013). Therefore, in synchronic phonemic systems and diachronic sound change, phonemes with large loads can form phonological rules, while those with small loads tend not to form rules and may be easily lost.[1] A single example suffices to demonstrate that phoneme structural load can explain the essence of numerous languages and dialects in synchronic systems and diachronic evolution.

[1] Research on Chinese dialect loads will be published separately.

IV. Conclusion

Based on the phoneme load algorithm proposed in this paper, we have calculated phoneme loads for initials, finals, and tones in Chinese, Tibetan, and Burmese languages, as well as in Chinese dialects and Tibetan dialects, revealing many patterns of linguistic phoneme structure, distribution, and function that differ from previous findings. We believe that through comparative load calculations, we can explore the status of particular phoneme combinations or individual phonemes within specific language systems. For Sino-Tibetan languages, this concerns the combinatorial patterns of initials, finals, and tones and their respective combinatorial capacities. Although we can examine the usage status of specific phonemes within language systems by counting their occurrence frequencies, this approach cannot truly describe or explain the phonological status of each initial/final/tone and their interrelationships.

In summary, this research takes the monosyllabic initial/final/tone phoneme system as its computational object, calculating opposition frequencies and load quantities for phoneme units. How to apply this method to calculate language load at the lexical, phrasal, and sentential levels requires further investigation, particularly regarding special phonetic phenomena in natural speech flow such as connected speech processes, coarticulation, and contextual phoneme interactions. Currently, contributions from these phenomena to language load cannot be quantified. Furthermore, how to further extend this computational method to applications in linguistic typology, historical linguistics, language acquisition, speech recognition, corpus design, and language model optimization also requires deeper research.

References

北京大学中国语言文学系语言学教研室编，王福堂修订. 2003. 汉语方音字汇 [*Chinese Dialect Character Compendium*]. Beijing: 语文出版社 [Language and Literature Press].

Bloomfield, Leonard. Translated by Yuan Jiahua, Zhao Shikai, Gan Shifu; edited by Qian Jinhua. 1980. 语言论 [*Language*]. Beijing: 商务印书馆 [Commercial Press].

Bloch, Bernard & G. L. Trager. Translated by Zhao Shikai. 1965. 语言分析纲要 [*Outline of Linguistic Analysis*]. Beijing: 商务印书馆 [Commercial Press].

Kong Jiangping. 2013. 音位负担计量研究——以藏缅语为例 [Quantitative Research on Phoneme Load: A Case Study of Tibeto-Burman Languages]. In 大江东去：王士元教授八十岁贺寿文集 [*Great River Flows East: Festschrift for Professor Wang Shiyuan' s 80th Birthday*], edited by Shi Feng & Peng Gang, 215-220. Hong Kong: 香港城市大学出版社 [City University of Hong Kong Press].

Sapir, Edward. Translated by Lu Zhuoyuan; edited by Lu Zhiwei. 1985. 语言论 [*Language*]. Beijing: 商务印书馆 [Commercial Press].

Chao Yuen-ren. Translated by Ye Feisheng. 1985. 音位标音法的多能性 [The Non-Uniqueness of Phonemic Solutions]. In 赵元任语言学论文选 [*Selected Linguistic Papers of Chao Yuen-ren*]. Beijing: 中国社会科学出版社 [China Social Sciences Press].

Mathesius, Vilem. 1929. "La Structure Phonologique du Lexique du Tchèque Moderne." *Travaux du Cercle Linguistique de Prague* 1: 67-84.

Jakobson, Roman. 1931. "Prinzipien der Historischen Phonologie." *Travaux du Cercle Linguistique de Prague* 4: 246-267.

Trubetzkoy, Nikolai. 1939. *Grundzüge der Phonologie. Travaux du Cercle Linguistique de Prague* 7.

Hockett, Charles F. 1955. *A Manual of Phonology* (= *International Journal of American Linguistics* 21). Indiana University Press.

Hockett, Charles F. 1967. "The Quantification of Functional Load." *Word* 23: 320–339.

Greenberg, H. H. 1959. "A Method of Measuring Functional Yield as Applied to Tone in African Languages." Georgetown University, Monograph Series on Languages and Linguistics 12: 7–16.

Hoenigswald, H. M. 1965. *Language Change and Linguistic Reconstruction*. Chicago: University of Chicago Press.

Wang, William S-Y. 1967. "The Measurement of Functional Load." *Phonetica* 16: 36–54.

King, R. D. 1965. *Functional Load: Its Measure and Its Role in Sound Change*. University of Wisconsin PhD dissertation.

King, Robert. 1967a. "Functional Load and Sound Change." *Language* 43.

King, Robert. 1967b. "A Measure for Functional Load." *Studia Linguistica* 21: 1–14.

Shannon, C. E. & W. Weaver. 1949. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Shannon, Claude E. 1951. "Prediction and Entropy of Printed English." *Bell Systems Technical Journal* 30: 50–64.

Kucera, Henry. 1963. "Entropy, Redundancy and Functional Load."In *American Contributions to the Fifth International Conference of Slavists*, 191–219. Sofia, Bulgaria. The Hague: Mouton.

Surendran, Dinoj & Partha Niyogi. 2003. "Measuring the Usefulness (Functional Load) of Phonological Contrasts." Technical Report TR-2003-12, Department of Computer Science, University of Chicago.

Surendran, Dinoj & Gina-Anne Levow. 2004. "The Functional Load of Tone in Mandarin is as High as that of Vowels." In *Proceedings of the International Conference on Speech Prosody 2004*.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.