

汉字同音率、同音度及同音字音节个数 随同音度增加而递减的规律

尹文刚

中国科学院心理研究所 北京 100101

提要 本研究提出同音率和同音度的概念,作为同音字特性的量的指标,并计算出汉字同音率为80.49%,汉字同音度的均值为7.85,且为偏态分布。通过对汉字同音度的研究,揭示出汉字同音音节个数随同音度增加而递减的规律。对此规律的讨论引申到可能的语言发展中的“清晰原则”。

关键词 同音字 同音率 同音度 平均同音度

中图分类号 H017 **文献标识码** A **文章编号** 1671-9484(2003)04-0003-04

1 前言

汉字的单音节性限制了汉字在语音维度上的发展,使其至今未能成为表音体系的文字,并且造成汉语发展中的一大基本矛盾:一方面汉字字数随着社会经济文化的发展而不断地增加,一方面汉字音节却又十分有限。这种矛盾的发展,导致了汉字中大量同音字的涌现,以致使现代汉语中同音字的大量存在成为汉语的一个主要特点。

作为一种重要的语言现象,特别对汉语来说是如此,同音字是一个很有意义的研究课题,对其不仅要有定性的描述,并应有定量的测量。自20世纪80年代以来,有关汉语语音的专题研究日渐增多,涉及到汉语语音的诸多方面,包括声母和辅音、韵母元音和复合元音、重音与轻声,其中尤以汉语声调的实验研究最为深入。但对汉语同音字的探讨有些不足,目前还未深入涉及到同音程度的数量化研究。本文在这里即提出同音率和同音度的概念以作为同音字特性的量的指标。

2 同音率 H_r

同音率定义为同音字音节在总音节中的比率。

设汉字字形 = G

设汉字音节 = P

设从 G 到 P 上的映射 = f

即: $f: G \rightarrow P$

其中 $D_f = G$, (D_f 系定义域)

[收稿日期] 2003年5月3日 [定稿日期] 2003年6月10日

$$Cf = P, (Cf \text{系值域})$$

如果存在若干字形 g_1, g_2, \dots, g_k , 且 $g_1 \neq g_2 \neq \dots \neq g_k$, 但 $f(g_1) = f(g_2) = \dots = f(g_k) = P_x$ 则 P_x 即为同音音节, 而 g_1, g_2, \dots, g_k 则是(互为)同音字。

又设 $P(H)$ 为同音音节的数目, 即:

$$P(H) = \sum_{i=1}^n P_{x_i}$$

则同音率为:

$$Hr = \frac{P(H)}{P}$$

其中, Hr 表示同音率, P 表示全部有意义音节, $P(H) \subset P$ 。

即是: 在 G 到 P 的映射中, 其同音音节的个数与总的映射个数之比就是同音率。

汉字同音率的值可以通过统计计算而求出。首先算出汉字全部有意义音节的个数, 即 P 。汉语音节是声母、韵母和声调的组合。设 A 为声母集合, 其元素有 21 个; 设 B 为韵母集合, 其元素有 39 个; 设 C 为声调集合, 其元素有 4 个; 则汉语音节可被看作是 $A \times B \times C$, 即集合 A 、 B 和 C 中元素的笛卡尔积。由于存在着声韵组合及声韵组合和声调配合的规则, $A \times B \times C$ 的实际值远远小于 $21 \times 39 \times 4$ 。通过统计计算, 实际存在的汉字全部有意义音节总数是 1307。

然后在这 1307 个有意义音节中找出同音音节。这一统计工作是在 1979 年版《新华字典》上进行的。统计结果, 共找出 1052 个同音音节。最后, 按前述同音率公式算出汉字同音率。其结果为:

$$\begin{aligned} Hr &= \frac{P(H)}{P} \\ &= \frac{1052}{1307} \\ &= 80.49\% \end{aligned}$$

3 同音度 Hd

同音度定义为同音音节的同音程度。在前述定义的 $G \rightarrow P$ 映射中, 若存在 g_1, g_2, \dots, g_k 且 $g_1 \neq g_2 \neq \dots \neq g_k$, 但 $f(g_1) = f(g_2) = \dots = f(g_k) = P_x$, 则:

$$Hd(P_x) = \sum_{i=1}^n g_i$$

即表示 P_x 的同音度。也就是, 某一同音音节的同音度等于与该同音音节相连的同音字个数。

例如, “bù” 是一个同音音节, 有 13 个汉字同“bù”相连(1979 年版《新华字典》), 它们是: 不, 吓, 钚, 布, 簿, 步, 埔, 埠, 部, 怖, 甬, 箬。故 $Hd(bù) = 13$ 。

上述定义的同音度给出了个别同音音节的同音程度, 为了得到对同音音节作为一个整体的同音程度的描述, 尚需要定义平均同音度。平均同音度由下述公式得出:

$$\overline{Hd} = \frac{\sum_{i=1}^n Hd(P_{x_i})}{N}$$

其中 \overline{Hd} 表示平均同音度, N 表示同音音节的个数, $\sum_{i=1}^n Hd(P_{x_i})$ 为所有同音音节同音度的总和。

汉字个别同音音节的同音度及平均同音度的值可通过统计计算而求出来。

统计是在 1979 年版《新华字典》上进行的。对 1052 个同音音节分别各个统计了相关联的汉字同音字数,从而求得各个同音音节的同音度。在此基础上,求出所有同音音节同音度的总和,即 8259。

按平均同音度公式,得出汉字平均同音度 = 7.85。

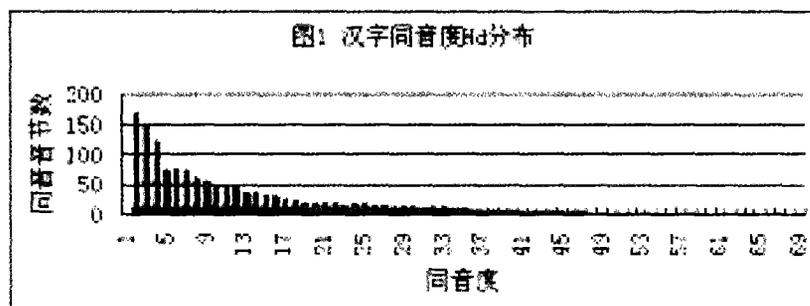
需要说明的是,同音度计算中对汉字的统计与字典中对汉字的编排有所不同。字典里收集了同一汉字的各种异体字,且把各个异体字亦算作不同的汉字而计入汉字字汇总数中。而这里在计算同音音节同音度时,为简明起见,未把异体汉字当作不同的汉字计入。

按上述公式算出的汉字平均同音度建立在众多的个别同音音节同音度数据的统计上,其标准差 $SD = 6.79$,最大值与最小值之间距为 67,最大值为 69,最小值为 2,在 2 与 69 之间分布着 40 个不同的同音度,众数为 2,中数为 6。这些同音度的分布呈一偏态状,其 skewness 值为 3.03, kurtosis 值为 15.81。

4 同音字音节个数随同音度增加而递减的规律

同音字的发展是否有规律可循? 由对汉字同音度的研究可以得到肯定的答案。

通过统计,可以把汉字同音度的分布状况求出来,如图 1 所示。



图中横坐标表示汉字同音度,纵坐标表示同音字音节个数。与各个同音度相对应的同音音节个数的分布呈现一条曲线,该曲线描绘了汉字同音度的分布情况,标示出具有各同音度的汉字同音音节的个数。从这个分布图上可以清楚地看到一个规律:汉字同音字音节的个数随着同音度的增加而递减。

这个规律展示出汉字同音字发展变化的模式,揭示出汉字同音字发展变化的一个规律。汉字同音字的发展变化不是任意的社会历史偶然过程,而是受制于特定的语言发展规则。这个规则调节着同音字的增长分布,在汉语音节体系中规定着同音负荷的比重。

同音字音节个数随同音度增加而递减的规律可能具有人们在认知心理上的原因。文字是人们创造的,是为人们交际而产生发展的,它的发展变化会自然受到人们在认知心理选择上的制约。很自然,文字的语音感知是越清晰越好,在人们对文字的使用中,这一认知心理上的倾向可能实际上起到了调节同音字分布的作用。文字同音负荷越重,语音感知越不清晰,自然会在历史演进中受到限制,从而使文字在语音上的定位社会历史性地遵从一种“清晰原则”,即文字在语音上的发展变化以保持相对的认知心理上的清晰为原则,进而使同音负荷合理分布。

同音字音节个数随同音度增加而递减的规律可能并不仅限于汉语,在其他语种中可能也存在,不过其表现形式会有所不同,大概它不会以音节为单位,而很可能是别的语音学结构。汉语由于其文字形式上的单音节性及其同音文字的大量存在,致使在汉语中存在着以音节为单位的明显的同音字音节个数随同音度增加而递减的规律。但是同音现象并不仅仅限于汉语,也并不仅仅限于字与音节的水平,它可

以存在于词的水平,也可以扩展到亚字水平上而表现为部分同音现象。因此,这一规律可以扩展成更为普遍的形式:同音单位个数随同音单位同音度的增加而递减。很有可能这是一个重要的语言发展的普遍规律,而其普遍性的基础则在于语言发展中可能存在的认知心理上的“清晰原则”。

参考文献

- 罗常培 王 均 1981 《普通语言学纲要》,北京:商务印书馆。
陈保亚 1999 《二十世纪中国语言研究方法论》,济南:山东教育出版社。
徐世荣 1980 《普通话语音知识》,北京:文字改革出版社。
王理嘉 1989 实验语音学与传统语音学,《语文建设》第1期,55-58页。
林 焘 1990 《语音探索集稿》,北京:北京语言学院出版社。

作者简介

尹文刚,男,1955年3月出生。中国科学院心理研究所研究员,博士生导师。1978年毕业于北京首都医科大学医疗系,后留校任生理学教师。1979年考入中国科大研究生院,1983年毕业获心理学硕士学位。1987年留学英国,1991年获神经心理学博士学位,后留校工作,主持过中、日、英三国失语症研究项目。1994年回国后在中国科学院心理研究所任研究员,继续从事神经语言学和认知神经心理学方面的研究工作。研究领域为神经语言学、心理语言学、认知神经心理学,主攻失语症和失读症,尤其是汉语表层失读和深层失读,发表论文多篇。

Homophone Ratio, Homophone Degree, and the Rule that the Number of Homophone Syllable Decreases as the Degree of Homophone Increases

Yin Wengang

Institute of Psychology, Chinese Academy of Sciences, Beijing 100101

Abstract Two concepts, homophone ratio (Hr) and homophone degree (Hd), are proposed here as indexes of quantitative measure of characteristic of homophone. Through calculation, the ratio of Chinese homophone is obtained which is 80.49%, the average degree of Chinese homophone is obtained which is 7.85, while the degree of Chinese homophone is bially distributed. By studying the degree of Chinese homophone, a rule is revealed that the number of homophone syllable decreases as the degree of homophone increases. The implication of this rule probably relates a possible regulation of development of language, “the principle of clearness”.

Keywords homophone homophone ratio (Hr) homophone degree (Hd) average degree of homophone