



JCLEE

Journal of Chinese Language Education and Evaluation

JCLEE, Vol. 2, No. 1, 2026, pp.84-97.

Print ISSN: 3078-283X; Online ISSN: 3104-5030

Journal homepage: <https://www.cleejournal.com>

DOI: <https://doi.org/10.64058/jclee260107qmccl>



《国际中文教育中文水平等级标准》各等级“汉字表”和“词汇表”数量匹配探讨

陶 炼 (Tao Lian)

摘要: 对六种不同高频词语表的词语与其构词用字的数量统计表明, 高频词语的数量和高频词语的构词用字数量之间存在稳定的关联; 而《国际中文教育中文水平等级标准》各个水平等级的“词汇表”和“汉字表”的词汇数量和汉字数量匹配则呈现出与之显著的差异, 其结果便是“词汇表”中出现较多使用频率偏低的词语。出现这一状况的根本原因在于“汉字表”和“词汇表”数量设置的不匹配, 而“创新”的“先汉字、后词汇”的“汉字表”“词汇表”编制方法则为这一不匹配的数量设置的通行铺平了道路。

关键词: 《国际中文教育中文水平等级标准》; 汉字表; 词汇表; 字表和词表的数量匹配; “先汉字、后词汇”

作者简介: 陶炼, 副教授, 复旦大学国际文化交流学院, 研究方向: 国际中文教育和现代汉语。电邮: taolian@fudan.edu.cn。

Title: On the Quantity Matching of “Chinese Characters List” and “Words List” in *Chinese Proficiency Grading Standards for International Chinese Language Education*

Abstract: Statistics on the quantity of words and the quantity of Chinese characters used for forming these words in six different high-frequency words lists shows that there is a stable correlation between the quantity of high-frequency words and the quantity of Chinese characters

for forming these words. The setting of the quantity of “Words List” and the quantity of “Chinese Characters List” in *Chinese Proficiency Grading Standards for International Chinese Language Education* shows a substantially different quantity correlation which causes a lot of comparatively low-frequency words entering its “Words Lists”. This is caused by the setting of mismatching quantity of “Chinese Characters List” and “Words List”, and the so called innovative “Chinese Character precedes Words” approach also paved the way for it.

Keywords: *Chinese Proficiency Grading Standards for International Chinese Language Education*; Chinese Characters List; Words List; Quantity matching of Chinese characters list and words list; Chinese Characters precedes Words

Author Biography: Tao Lian, Associate professor; ICES, Fudan University. Research Interests: International Chinese language education and Modern Chinese language.

E-mail: taolian@fudan.edu.cn .

一、分等分级“汉字表”和“词汇表”的数量设置和数量匹配问题

编制分等分级的“汉字表”和“词汇表”都将面对无法回避的各等级“汉字表”和“词汇表”的数量设置问题，以及同一等级的“汉字表”和“词汇表”的数量匹配问题，要给这两个问题给出一个能够令人信服的答案却绝非易事。《汉语水平词汇与汉字等级大纲》（国家对外汉语教学领导小组办公室汉语水平考试部 1992）从以下两个方面来回答这一问题：一是高频词语的覆盖率，二是“对外汉语教学几十年的实践经验”。在刘英林、宋绍周（1992）看来，3000、5000、8000 高频词语的覆盖面分别为 86%、91%和 95%，因而这三个数字成为乙级、丙级和丁级词汇总量的基准；另一方面，在教学实践中，一年级前一阶段的词汇量接近 1000，因而 1000 成为甲级词的数量指标，而一年级和二年级结束时，词汇量分别在 2000-4000 和 5000 左右，于是 3000 和 5000 成了乙级和丙级词汇总量的基准。然而这样的论证，说服力却有限。李清华（1999, p. 56）就提问“为什么是 95%而不是 96%或 97%”？据她的了解，以 95%为标准“也没有什么道理和根据，好像只是因为 95%相对于 96%和 97%来说是个逢五的整数。”教学实践中的差别更是显而易见。李清华认为“词汇等级大纲”的总量应为 15000 左右，姜德梧（2004, p. 88）则认为应在 10000-12000 左右。他们论证的方法各不相同，但说服力都同样有限。

在论及《汉语国际教育用音节汉字词汇等级划分》（中华人民共和国教育部、国家语言文字工作委员会 2010。以下简称《等级划分》）各等级的“汉字表”和“词汇表”的数量设置依据时，刘英林、马箭飞（2010, p. III）说：“从‘汉字覆盖率’‘音节的统计与覆盖率’和‘与未来发展接轨’三个角度综合考虑”，一级汉字表的基本量“定为 900 字为好”。具体来说就是 900 字能够覆盖“大型对话体动态语料”的 92.10%，覆盖《基本音节表》50%以上的音节，是即将研制的“自动化汉语口语水平考试（HKK）C 等”300 个汉字的 1/3。显然，李清华的提问同样可以运用于此：要

是我们把一级汉字的数量调整为 850 或 950 字，三项覆盖率也相应地变低或变高一点儿，会有什么根本的不妥吗？

刘英林、马箭飞（2010，p. XI）在同一篇文章中还有如下两段相关陈述：

根据汉语国际教育的总体目标和 60 年对外汉语教学、世界汉语教学的经验，《等级划分》三级水平的汉字基本量定为 2700 字。……将 2700 字均衡地分为三个等级，一级汉字 900 字；二级汉字 900 字……；三级汉字 900 字。

根据世界各地国际汉语教学和国内对外汉语教学的总体战略目标，《等级划分》三级水平的词汇基本量定为 9600 个左右。

这里倚重的是“教学的经验”。但教学经验显然存在着不小的差异，对于具有不同经验的人来说，倚重“经验”的论证说服力就显得弱了一些。

相比于各个等级“汉字表”和“词汇表”的容量设置，同一等级“汉字表”和“词汇表”之间的数量匹配是一个更少得到关注的问题。

我们可以看到，迄今为止国际中文教育领域推出的多种分等分级“汉字表”和“词汇表”，等级划分框架和各个等级“汉字表”“词汇表”的数量大都各有考量，差别明显，甚而被看作是一个创新的所在；而相关的论说却大多简要，只是表明某些关注，很难说是全面的论证，因而迄今为止也不存在可以目之为“共识”的分等分级“汉字表”和“词汇表”的架构。由表 1 列举的五种“大纲”分等分级架构的各种不同就可以看出些许端倪。

《汉语水平词汇与汉字等级大纲》		甲级	乙级	丙级	丁级
	汉字	800	1604	2205	2875
	词汇	1033	3051	5253	8822
《高等学校外国留学生汉语言专业教学大纲》		一年级		二年级	三、四年级
		一级	二级		
	汉字	795	1491	2036	2503
	词汇	993	2704	4919	7554
《高等学校外国留学生汉语教学大纲·长期进修》		初等	中等	高等	
	汉字	1411	2126	2629	
	词汇	2399	5249	8009	

《国际汉语教学通用课程大纲》		一级		二级		三级		四级		五级	
	汉字	150		300		450		600		800	
	词汇	300		600		900		1200		1500	
《汉语国际教育用音节汉字词汇等级划分》		初等			中等			高等			
		一级	二级	三级	四级	五级	六级	高级	高级附录		
	汉字	300	900		1800			2700	3000		
	词汇	505	1342	2245	5456			9631	11092		

表 1：五种国际中文教育大纲的水平等级框架及各等级累计汉字、词汇数量

Table 1: The frameworks and the cumulative numbers of Chinese characters and words at each level of five syllabuses for international Chinese language education

二、《国际中文教育中文水平等级标准》“汉字表”和“词汇表”的数量匹配

《国际中文教育中文水平等级标准》（中华人民共和国教育部、国家语言文字工作委员会发布 2021。以下简称《等级标准》）“汉字表”和“词汇表”承袭了（刘英林等 2021，p. 5）《等级划分》“汉字表”和“词汇表”的初等、中等、高等三级框架以及数量指标，只是对三个等级内部做了等级的下位细分或合并；还调整了部分字、词以及字、词的等级归属。（刘英林等 2021，p. III-IV）两者的等级框架及各个水平等级的字、词累计数量对应关系见表 2。

		初等			中等			高等	
		一级	二级	三级	四级	五级	六级	高级	高级附录
《等级划分》	汉字	300	900		1800			2700	3000
	词汇	505	1342	2245	5456			9631	11092
《等级标准》	汉字	300	600	900	1200	1500	1800	3000	
	词汇	500	1272	2245	3245	4316	5456	11092	

表 2：《等级划分》和《等级标准》“汉字表”“词汇表”水平等级框架及各等级累计字、词数量对比

Table 2: The frameworks and the cumulative numbers of Chinese characters and words at each level of Graded Classification and Grading Standards

《等级标准》“使用说明”说，参考多种资料，“根据汉字的流通度、构词能力、书写难易度、文化内涵等因素，将初等和中等 1800 个汉字均分到一一六级，每级 300 字”；“将初等和中等 5456 个词语根据词汇难度及使用频率，细分为一一六级”，每级数量不等。（刘英林等 2021, p. III-IV）至于汉字流通度、词汇难易度这些标准怎样具体影响或决定了一个汉字或词语的等级归属？为什么汉字等级可以均分词汇却没有均分？诸如此类影响着“汉字表”和“词汇表”的数量设定和数量匹配的问题，《等级标准》的编制者都没有再进一步的说明。

三、高频词语表中的汉字和词语数量匹配

以频率词表作为编制分等分级“汉字表”和“词汇表”的主要依据是国内国际中文教育界的传统。《汉语水平词汇与汉字等级大纲》“的主要依据是国内影响较大的七个动态性频度统计词典、词表与字表”。（刘英林、宋绍周，1992, p. 5）《等级划分》的编制也延续着这一传统，“汉字表和音节表的筛选必须以当代大型动态语料库和多种具有代表性的字表、词表为依据”，（刘英林、马箭飞 2010, p. III）词汇表的筛选也一样；而这其中字表、词表的作用主要在于“有据可依”地“查漏补缺”，（刘英林、马箭飞 2010, p. IV）也就是说，是起辅助作用的，起主导作用的还是字词的大型动态频率调查结果。这就决定了《等级划分》以及《等级标准》中的“词汇表”和“汉字表”从本质上来说是一种频率字表或词表。因而破解分等分级“词汇表”和“汉字表”数量匹配问题的关键或许就藏在频率词表和字表之中。

《等级划分》依据的当代大型动态语料有五种：（1）《广播电视对话节目语料》基本数据（2005-2008）；（2）《有声媒体语料》基本数据（2005-2008）；（3）《报纸、广播电视、网络新闻综合语料》基本数据（2005-2008）；（4）《九年义务教育（新课标）语文科语料》基本数据；（5）《国家语委现代汉语语料库（平衡语料库）》基本数据。此外还有十种现代汉语词典、等级大纲、常用字表和词表。（刘英林、马箭飞 2010, p. II）《等级标准》在修订“汉字表”和“词汇表”的时候，又参考了其他若干种材料。（刘英林等 2021, p. III）

我们没有能力获得与《等级划分》或《等级标准》相同的依据材料，我们用的是以下三种频率词表：（1）《媒体高频词语表》（2010-2020）；（2）《国家语委现代汉语语料库词频表》；（3）《现代汉语常用词表（第 2 版）》。这三种材料，《媒体高频词语表》与《等级划分》的前三项语料源于同一个“年度媒体用字用语调查”项目；《国家语委现代汉语语料库词频表》也同样源于国家语委现代汉语语料库；《现代汉语常用词表》也在《等级划分》的参考词表之列，只不过入列的不是“第 2 版”，而是“草案”版。这就是说，我们所用的这三种材料，与《等级划分》和《等级标准》所依据的材料十分相近，因而所得的结果对于认识《等级划分》和《等级标准》应该具有相当高的参考价值。

我们从《媒体高频词语表》（2010-2020）中提取出四个表：分别是 2010 年、2015 年和 2020 年

三个年度《媒体高频词语表》，以及把 2010—2018 年这 9 年的《媒体高频词语表》整合而成的一个《2010—2018 年度媒体高频词语表》，表中的词语按出现年数由多到少、平均序号由少到多排列。加上《国家语委现代汉语语料库词频表》和《现代汉语常用词表（第 2 版）》，一共是 6 个高频词语表。我们考察的重点是高频词语的数量和所用汉字的数量之间是否存在相当稳定的关联，因此，我们从这 6 个词表中，由高频向低频，即从头到尾，截取与《等级标准》各个水平等级相同数量的词语，统计它们所用汉字的数量。统计用教育部语言文字应用研究所计算语言学研究室的“字词频统计”软件完成。统计结果见表 3。

各个水平等级词汇的累计用字量	《等级标准》一级至七-九级累计词汇量						
	一级 500	二 级 1272	三 级 2245	四 级 3245	五 级 4316	六 级 5456	七-九级 11092
“2010 年度媒体高频词语表”	445	818	1130	1384	1594	1806	2564
“2015 年度媒体高频词语表”	445	815	1116	1360	1581	1774	2543
“2020 年度媒体高频词语表”	464	816	1099	1310	1481	1683	2443
“2010—2018 年度媒体高频词语表”	443	807	1103	1354	1582	1786	2558
《国家语委现代汉语语料库词频表》	410	816	1154	1431	1702	1943	2820
《现代汉语常用词表（第 2 版）》	440	844	1183	1471	1753	1991	2850
《等级标准》“词汇表”	300	600	900	1200	1500	1800	3000

表 3：与《等级标准》各个水平等级数量相同的高频词语的累计用字数量

Table 3: The cumulative numbers of characters used by the high-frequency words according to the frameworks of *Grading Standards*

表 3 的数据表明，6 个高频词语表各个水平等级词汇的用字数量波动的幅度不大，这就意味着高频词语的数量跟它们的构词用字数量之间存在着相当稳定的关联；当然，也有一定的上下波动，比如《国家语委现代汉语语料库词频表》和《现代汉语常用词表（第 2 版）》从第二或第三级开始呈现出用字量逐渐增大的趋势，最终比“2010—2018 年度媒体高频词语表”的用字多出 300 左右。我们的这一个比对发现，虽然仍然不能够给“每一级的汉字和词汇为什么是这个数量”这一问题一个客观的答案，但为以频率为主要依据的分等分级“汉字表”和“词汇表”的数量匹配问题，找到了一个颇为“客观”的设置标准和评价依据。既然不同水平等级的高频汉字和词语之间存在着相对稳定的数量匹配关系，而《等级标准》也是以汉字和词语的频率属性作为等级划分的最主要依据，那么《等级标准》各个水平等级的“汉字表”和“词汇表”之间的数量匹配也应当反映高频词表中的汉字和词语的数量匹配关系。

四、《等级标准》各个等级“汉字表”和“词汇表”数量匹配的偏差

与表 3 所示频率词表的各个水平等级词汇的用字数量相比,《等级标准》各个水平等级词汇的用字数量呈现出显著的差异。具体表现为:一级到六级,相对于“词汇表”的数量设定,“汉字表”的数量明显过低,低了 100-200 多字;七-九级,“汉字表”的数量明显过高,跟前 4 个表比高出 450 字左右,跟后边两个数量较大的表比,也还是高出 150 字或更多一些。图 1 是《等级标准》从一级到七-九级词汇的用字量跟《2010-2018 年度媒体高频词语表》《国家语委现代汉语语料库词频表》和《现代汉语常用词表(第 2 版)》三个高频词表相应数量词语平均用字量的对比,更加直观地显示出两者的差别。

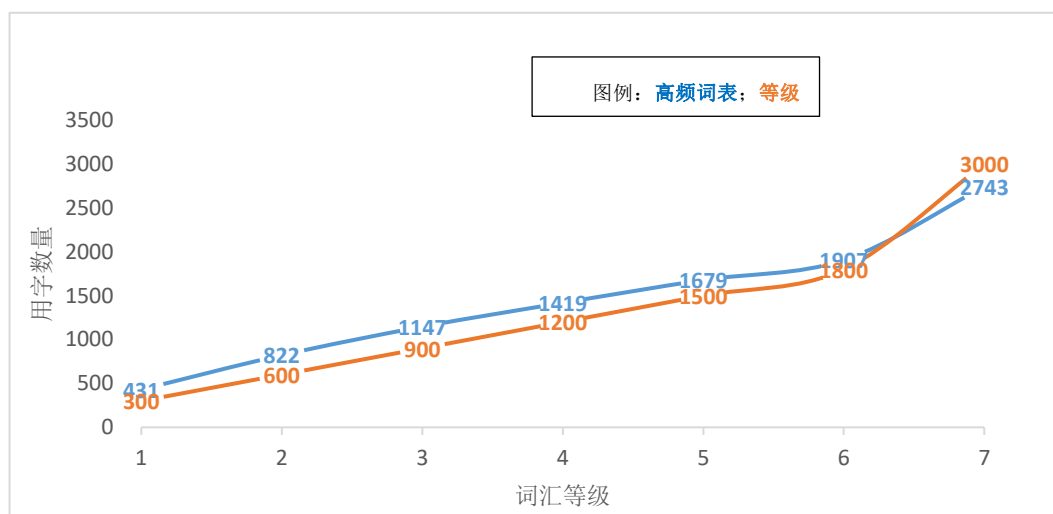


图 1:《等级标准》词语用字量与三个高频词语表词语平均用字量对比图

Figure 1: Comparison of the Number of Characters Used in *Grading Standards* and the Average Number of Characters Used in Three High-Frequency Words Lists

《等级标准》各个水平等级词汇的用字数量跟频率词表词语的用字数量差异显著,说明《等级标准》各个水平等级的字量和词量的设置没能反映出高频词语内在的词、字数量的客观匹配特征。

《等级标准》“汉字表”和“词汇表”数量匹配呈现出来的偏差,对于《等级标准》“词汇表”的编制也将带来不利影响。

五、《等级标准》“汉字表”和“词汇表”数量匹配偏差的影响

由于一级到六级《等级标准》的问题是相对于“词汇表”的数量设定,“汉字表”的汉字数量偏少,而七-九级的问题是“汉字表”的汉字数量偏多,我们就以一级和七到九级两个端点为例来探讨这种数量匹配偏差可能会造成的影响。

先看初等一级。高频词语表的前 500 个词语,用字量在 430 个左右,而《等级标准》一级词汇 500 个词语,仅用汉字 300 个,是 500 字的 60%。可以少学 40% 的汉字,这看起来减少了汉字学习的负担,有利于初等一级阶段的汉字学习,但问题是仅用 300 个汉字,不可能在高频词语表的最前端,

即最高频段找到 500 个仅用这 300 个汉字构成的最高频词语，因为那需要 430 个左右的汉字；由这 130 个左右汉字构成的最高频词语，由于这些汉字不在一级“汉字表”内而不能入选一级“词汇表”；而要填满这些空位、选出 500 个词语，就不得不向排位在前 500 位之后的、频率相对较低的词语中去选择用这 300 个汉字构成的词语。这就是说，仅用 300 个汉字，会使筛选出来的 500 个词语中，包含较多位于高频词表前 500 位之后的、也就是使用频率相对较低的词语。这就降低了一级“词汇表”的高频词语属性，因而也降低了一级词语的学习价值和使用价值。

我们依据“2010-2018 媒体高频词语表”、《国家语委现代汉语语料库词频表》和《现代汉语常用词表（第 2 版）》模拟构建一个 500 词的一级词汇表。如果从这三个表中各取频率最高的前 550 个词，合并成候选词语表，得到不同的词语 784 个，其中三个或两个表共有的词语是 526 个，取其中排位最靠前的 500 个构成模拟词表，那么，用到的汉字是 426 个，三个高频词表最靠前的 500 词语分别有 376、424、421 个入选。如果从这三个表中各取频率最高的前 1000 个词，合并成候选词语表，得到不同的词语 1482 个，其中三个表共有的词语是 578 个，取其中排位最靠前的 500 个构成模拟词表，那么，用到的汉字是 442 个，三个高频词表最靠前的 500 词语分别有 368、380、410 个入选。以上模拟操作表明，即便采用同样的词语筛选标准和操作流程，候选词语表取词范围大小的不同，也会影响最终高频词语的入选数量。要多入更多选高频词语，就要尽量缩小取词范围。我们曾把取词范围定为前 500，结果三个或两个表共有的词语是 468 个，不能满足一级词汇表 500 的数量要求，所以我们把取词范围放宽到前 550 个。

《等级划分》一级词语的筛选，是从 5 个大型频率统计词表的前 1600 个词中筛选候选词语。之所以要划定超过三倍多数量的选词范围，我们认为或许是因为较小的选词范围里没有足够多的全部用一级汉字构成的词语；但划定一个较大的选词范围，就给频率较低的词语入选一级词表提供了机会；事实上一级词表中还包含少量来自词表前 1601-3000 位次的词语。（刘英林、马箭飞 2010, p. IV）如果用我们的三个高频词表来衡量，《等级标准》“一级词语”的 500 个词出现在这三个高频词表的四个不同排列序号区间的数量见表 4，其中处于词表前 1-500 范围的数量只有 145 或 167 个，跟我们的模拟一级词汇表相比，《等级标准》“一级词汇表”的 500 词中，处于最靠前 500 位的高频词语太少，而频率偏低的词语太多。

三个高频词语表的序号区间	1-500	501-1000	1001-5000	5001 以上	总计
2010-2018 媒体高频词语表	145	62	184	109	500
国家语委现代汉语语料库词表	167	69	161	103	500
现代汉语常用词表（第 2 版）	167	77	134	122	500

表 4：《等级标准》一级“词汇表”500 个词语在三个高频词语表中的分布情况

Table 4: Distribution of the 500 words of Level 1 from *Grading Standards* in three high-frequency words lists

造成《等级标准》一级“词汇表”中频率偏低的词语占比过高,在我们看来,原因有二:一是设定的汉字表、词汇表数量匹配不合适,汉字数量相对过少,二是选词范围设定过大,这都给频率偏低的词语进入“词汇表”创造了机会。

如果维持一级“词汇表”500词汇量不变,那么应该将汉字量增加到440上下,以保障入选词语有较高的使用频率,同时也为汉字的学习和使用提供更多的机会;如果维持一级“汉字表”300汉字量不变,那么应该将词汇量减少到300个左右,同样是为了保障入选词语有较高的使用频率,同时也维持较低的汉字学习量。我们的三个高频词语表前300词的用字量分别是308、275和286。

除了初等一级,《等级标准》初等二级到中等六级,同样存在着程度不等的汉字数量相对不足的问题。在我们看来,这同上面探讨的初等一级相似,也都会迫使“词汇表”去吸纳使用频率相对偏低的词语。

再看高等七—九级。《等级标准》“词汇表”11092个词语,用了2971个汉字,(另有29个汉字是常用的地名和姓氏用字,实际上并没有出现在“词汇表”中。)但我们考察的三个高频词语表的前11092个高频词语的用字数量分别是2558、2820和2850,并没有2971那么多,差值在121—413之间。要多少高频词语才用得到3000个汉字?《2010—2018年度高频词语表》前16000个词语用字3014个;用字较多的《国家语委现代汉语语料库词频表》前12500个词语用字2982个,《现代汉语常用词表(第2版)》前12000个词语用字2942个。所以,为了实现出现3000个汉字的目标,就必须加入使用频率实际不够高的汉字,以及由这些汉字构成的、出现在高频词语表11092位之后的、使用频率实际不够高的词语。这就使得“汉字表”和“词汇表”中都出现了使用频率偏低的、本不该出现的汉字和词语,自然给汉字和词汇的学习增添了负担,降低了效用;而低频词语的入选,也意味着有同样数量的原本可以凭借使用频率入选的词语被挤了出去,这又从另一方面削弱了“词汇表”的高频属性。

考察《等级标准》“词汇表”中包含仅构词1个的汉字的词语在我们的三个高频词语表中的位次,可以发现,在三个或两个高频词语表中处于第11092位以后的词语中所包含的仅构词1个的汉字有以下302个:

藹、隘、芭、靶、掰、柏、斑、瓣、谤、磅、镑、煲、豹、毙、痹、眨、飙、缤、秉、哺、眯、沧、蹭、岔、掺、搀、馋、昌、猖、嫦、橙、秤、弛、侈、橱、踰、炊、捶、锤、醇、鹵、葱、簇、翠、捣、涤、笛、蒂、巅、刁、叨、谍、碟、眈、炖、哆、舵、讹、娥、厄、鳄、餌、芳、绯、诽、芬、敷、斧、俯、丐、竿、冈、膏、耿、躬、沽、菇、卦、龟、骇、焊、撼、浩、禾、阍、虹、吼、弧、唬、猾、槐、恍、讳、霍、讥、缉、畸、脊、颊、歼、柬、贱、溅、浆、桨、跤、蕉、礁、狡、绞、矫、荆、窘、灸、鞠、橘、沮、炬、锯、爵、钧、骏、凯、楷、苛、垦、吭、扼、窟、垮、拷、

旷、窥、腊、婪、澜、唠、涝、磊、蕾、棱、敛、潦、陵、遛、窿、卤、挛、抡、螺、莽、茅、昧、
 觅、泌、緬、渺、蔑、谬、蘑、沐、睦、暮、穆、呐、馁、尼、黏、畔、袍、沛、抨、烹、鹏、媲、
 萍、仆、沭、栖、钳、乔、俏、翘、撬、怯、禽、擎、丘、囚、裙、绒、冗、瑞、臊、鲨、膳、瞻、
 捎、勺、慑、矢、侍、嗜、赎、曙、栓、涮、酥、溯、蒜、嗦、潭、涛、剃、涕、舔、秃、颓、唾、
 蛙、婉、汪、妄、纬、瘟、紊、翁、涡、沃、巫、呜、锡、熄、嬉、侠、霞、馅、腺、萧、淆、啸、
 屑、衅、腥、汹、锈、旭、恤、酗、絮、炫、靴、勋、驯、阄、衍、殃、秧、痒、漾、妖、椰、夷、
 怡、屹、役、弈、逸、荫、殷、莹、佣、咏、佑、渊、陨、攢、赃、枣、藻、皂、榨、瞻、蘸、爪、
 沼、肇、辙、贞、峙、窒、拄、贮、拙、灼、浊、酌、兹、棕、揍、卒、佐

在考察过程中，我们也注意到一些构词 2 个的汉字。例如：“荆棘”在高频词语表中的位置偏后，“棘”构成的另外一个词“棘手”位置也偏后；“沐浴露”位置偏后，“浴”构成的另一个词“浴室”位置也偏后；“沼泽”位置偏后，“泽”构成的另一个词“光泽”位置也偏后；“叨”构成的两个词“絮叨”“唠叨”位置都偏后；因此，“棘、浴、泽、叨”也都可以加入到上面的表中。这种情况我们并没有全面探查，这就意味着上面这个“低频”汉字表还有增加的可能。另有一种情况，“词汇表”中所列的词语位置在 11092 之外，但用那个“字”构成的其他词语却有资格出现在“词汇表”中，这样的汉字我们都没有放在上面的表里。例如《等级标准》“词汇表”中：用“驰”构成的唯一词语“驰名”在三个高频词语表中都位置偏后，但“奔驰”在三个高频词语表中都出现在 11092 位之前，所以尽管“词汇表”里没有“奔驰”，但我们也没有把“驰”列在上面的表中。

以上我们只是列出考察的结果，说明这些词语及其所用的某一个或所有构词汉字，从使用频率角度来看，没有资格出现在容量为 11092 词的“词汇表”中，同样也不应该出现在“汉字表”中。但这并不意味着这些字和词就一定不能够进入“汉字表”和“词汇表”。它们还可以依据来自其他方面的“有据可依”的“查漏补缺”进入“汉字表”和“词汇表”。比如“（牙）膏、（香）蕉、橘（子）、裙（子）、勺、痒、（肥）皂”，再比如“芳、芬、萍”，都是可以考虑加入的例子，但“查漏补缺”不该有很大的数量吧。

六、《等级标准》“汉字表”和“词汇表”数量匹配偏差原因探讨

《等级标准》“汉字表”和“词汇表”之所以表现出各个等级数量匹配的偏差，根本的原因在于没有认识到高频词语和它们的用字之间存在稳定的数量关联，因而给各个等级的“汉字表”和“词汇表”设定了并不互相匹配的数量指标。一级到六级，表现为“汉字表”相对于“词汇表”容量过小，汉字不能够满足高频词语的构词需求，于是不得不放弃带“生字”的高频词语，而去选择

使用频率较低的由“汉字表”中的字构成的词语；七一九级，表现为“汉字表”相对于“词汇表”容量过大，部分汉字并不用于相应频率区间的词语，于是为了用上这些汉字，不得不去使用频率较低的区间挑选使用这些汉字的词语，并把它们纳入“词汇表”，为此也不得不放弃同样数量的高频词语来腾出空位。因而无论过与不足，数量不匹配的“汉字表”和“词汇表”都使得编制出来的“词汇表”不得不容纳较多的使用频率偏低的词语，降低我们本当追求的“词汇表”的高频属性。

《等级标准》最终之所以接受并不匹配的各等级“汉字表”和“词汇表”数量设定，还与《等级划分》和《等级标准》所践行的“创新”的“先汉字、后词汇”的“汉字表”“词汇表”编制“新路向”密切相关。如果采用传统的“先词汇、后汉字”的编制方法，由于先编制“词汇表”后编制“汉字表”，“词汇表”的容量决定了依它而生的“汉字表”的容量，“汉字表”没有增减自身数量的自由，因而走“先词汇、后汉字”的老路，无论是否自觉意识到高频词语数量和它们的用字数量之间存在稳定的关联，都不可能出现“汉字表”和“词汇表”两者数量匹配失衡的问题。而“创新”的“先汉字、后词汇”“新路向”，先编制“汉字表”，后编制“词汇表”，就使得在设定了“汉字表”的容量之后，再去设定“词汇表”的匹配容量时，失去了来自先编制“词汇表”带来的程序性容量限定。这时候，如果没有对“汉字表”和“词汇表”存在客观的数量匹配的认知，各种实际上并不匹配的数量设定就会以各种看似“客观”的理由或是主观的“经验”为依据被确定下来。所以，是创新的“先汉字、后词汇”的“汉字表”“词汇表”编制“新路向”为各个水平等级本不匹配的“汉字表”“词汇表”数量设定的通行铺平了道路。

其实，刘英林、马箭飞已经触及到了高频词语和它们的构词用字之间的数量匹配问题。在刘英林、马箭飞的文章中，我们读到：“据我们统计，广播电视对话节目词频统计前 1600 词里出现了 900 字种，……前 5238 个词是由 1800 个字种组成的，……前 12112 词包含了 2700 字种”。（刘英林、马箭飞 2010, p. IV-V）正是他们的这一统计结果引导我们去讶异《等级划分》和《等级标准》的“词汇表”和“汉字表”的数量匹配关系为什么跟这个统计有这么大的不同，去探索高频词语的数量和它们的构词用字数量之间是不是存在着相当稳定的固有联系，这才最终有了这篇不甚成熟的探讨文字。这是我们要向刘英林、马箭飞两位表达由衷感谢的地方。而他们两位似乎并没有去深究他们的发现所蕴含的意义，只是把它看作划定选词范围的一个依据，仍旧一心奔赴“创新”的“先汉字、后词语”的“汉字表”“词汇表”编制“新路向”。只可惜最终《等级划分》和《等级标准》各个水平等级“汉字表”和“词汇表”的数量匹配，在我们看来，出现了偏差。

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

ORCID

Tao Lian ^{ID} <https://orcid.org/0009-0001-6344-611X>

References

国家对外汉语教学领导小组办公室（2002）：《高等学校外国留学生汉语教学大纲·长期进修》。北京语言文化大学出版社。

[Office of the Leading Group for Teaching Chinese as a Foreign Language (2002). *Syllabus for Teaching Chinese to International Students in Higher Education Institutions: Long-term Study*. Beijing Language and Culture University Press.]

国家对外汉语教学领导小组办公室（2002）：《高等学校外国留学生汉语言专业教学大纲》。北京语言文化大学出版社。

[Office of the National Leading Group for Teaching Chinese as a Foreign Language (2002). *Teaching Syllabus for Chinese Language Major for Foreign Students in Higher Education Institutions*. Beijing Language and Culture University Press.]

国家对外汉语教学领导小组办公室汉语水平考试部（1992）：《汉语水平词汇与汉字等级大纲》。北京语言学院出版社。

[Chinese Proficiency Test Division of the Office of the Leading Group for Teaching Chinese as a Foreign Language (1992). *The Syllabus of Graded Words and Characters for Chinese Proficiency*. Beijing Language and Culture University Press.]

国家汉办、教育部社科司、《汉语国际教育用音节汉字词汇等级划分》课题组（2010）：《汉语国际教育用音节汉字词汇等级划分（国家标准·应用解读本）》。北京语言大学出版社。

[Office of the National Leading Group for Teaching Chinese as a Foreign Language, Department of Social Sciences of the Ministry of Education, Project Team of *The Graded Chinese Syllables, Characters and Words for the Application of Teaching Chinese to the Speakers of Other Languages* (2010). *The Graded Chinese Syllables, Characters and Words for the Application of Teaching Chinese to the Speakers of Other Languages: National Standards Application Interpretation Edition*. Beijing Language and Culture University Press.]

国家汉语国际推广领导小组办公室（2008）：《国际汉语教学通用课程大纲》。外语教学与研究出版社。

[Office of the National Leading Group for Teaching Chinese as a Foreign Language (2008). *International Curriculum for Chinese Language Teaching*. Foreign Language Teaching and Research Press.]

姜德梧（2004）：“关于《汉语水平词汇与汉字等级大纲》的思考”，《世界汉语教学》（01）：80-89。

[Jiang Dewu (2004). “Some thoughts on *The Syllabus of Graded Words and Characters for Chinese Proficiency*.” *Chinese Teaching in the World* (01):80-89.]

教育部语言文字信息管理司（2011-2020）：《中国语言生活状况报告（2011-2020）》。商务印书馆。

[Department of Language Information Management of the Ministry of Education (2011-2020). *Reports on the Status of Language Life in China (2011 - 2020)*. Commercial Press.]

教育部语言文字信息管理司组编、李行健、苏新春主编（2021）：《现代汉语常用词表（第2版）》。商务印书馆。

[Department of Language Information Management of the Ministry of Education, Li Xingjian & Su Xinchun (Eds.) (2021). *List of Commonly Used Words in Modern Chinese (2nd Edition)*. Commercial Press.]

教育部中外语言交流合作中心、刘英林、马箭飞、赵国成主编（2021）：《国际中文教育中文水平等级标准（国家标准·应用解读本）》。北京语言大学出版社。

[Center for Language Education and Cooperation of the Ministry of Education, Liu Yinglin, Ma Jianfei & Zhao Guocheng (Eds.) (2021). *Chinese Proficiency Grading Standards for International Chinese Language Education: National Standard · Applied Interpretation Edition*. Beijing Language and Culture University Press.]

李清华（1999）：“《汉语水平词汇与汉字等级大纲》的词汇量问题”，《语言教学与研究》（01）：50-59。

[Li Qinghua (1999). “On the Quantity of Vocabulary of *The Syllabus of Graded Words and Characters for Chinese Proficiency*.” *Language Teaching and Linguistic Studies* (01):50-59.]

刘英林（2021）：“《国际中文教育中文水平等级标准》的研制与应用”，《国际汉语教学研究》（02）：6-8。

[Liu Yinglin (2021). “The Development and Application of *Chinese Proficiency Grading Standards for International Chinese Language Education*.” *Journal of International Chinese Teaching* (01):6-8.]

刘英林、李佩泽、李亚男（2021）：“汉语国际教育汉语水平等级标准全球化之路”。教育部中外语言交流合作中心主编：《国际中文教育中文水平等级标准（国家标准·应用解读本）》，北京语言大学出版社：1-10。

[Liu Yinglin, Li Peize, Li Yanan (2021). “A Path to the Globalization of Chinese Proficiency Standards in International Chinese Education”, In Center for Language Education and Cooperation (Eds.), *Chinese Proficiency Grading Standards for International Chinese Language Education: National Standard · Applied Interpretation Edition*, Beijing Language and Culture University Press:1-10.]

刘英林、马箭飞（2010）：“研制《汉语国际教育用音节汉字词汇等级划分》探寻汉语国际教育新思维”，《汉语国际教育用音节汉字词汇等级划分（国家标准·应用解读本）》，北京语言大学出版社。

[Liu Yinglin & Ma Jianfei (2010). “The Development of *The Graded Syllables, Characters and Words*, Exploring the new prospective of global Chinese education”, *The Graded Chinese Syllables, Characters and Words for the Application of Teaching Chinese to the Speakers of Other Languages: National Standards · Application Interpretation Edition*, Beijing Language and Culture University Press.]

刘英林、马箭飞（2012）：“再论汉语国际教育新思维——解读和应用《等级划分》的若干问题”，《第十届国际汉语教学研讨会论文选》，万卷出版公司：405-417。

[Liu Yinglin & Ma Jianfei (2012). “Revisiting New Approaches to International Chinese Education: Some Issues in Interpretation and Application of *Graded Classification*”, *Proceedings of the 10th International Conference on Chinese Language Teaching*, Wanjuan Publishing Company :405-417.]

刘英林、宋绍周（1992）：“论汉语教学字词的统计与分级（代序）”，《汉语水平词汇与汉字等级大纲》，北京语言学院出版社：1-22。

[Liu Yinglin & Song Shaozhou (1992). “On the Statistics and Grading of Chinese Characters and Words for Chinese Language Teaching (Preface on Behalf)”, *The Syllabus of Graded Words and Characters for Chinese Proficiency*, Beijing Language and Culture University Press:1-22.]

中华人民共和国教育部、国家语言文字工作委员会（2010）：《语言文字规范 GF0015-2010 汉语国际教育用音节汉字词汇等级划分》。北京语言大学出版社。

[Ministry of Education of the People's Republic of China & National Language and Character Work Committee (2010). *The Graded Chinese Syllables, Characters and Words for the Application of Teaching Chinese to the Speakers of Other Languages*. Beijing Language and Culture University Press.]

中华人民共和国教育部、国家语言文字工作委员会（2021）：《语言文字规范 GF0025-2021 国际中文教育中文水平等级标准》。北京语言大学出版社。

[Ministry of Education of the People's Republic of China & National Language and Character Work Committee (2021). *Chinese Proficiency Grading Standards for International Chinese Language Education*. Beijing Language and Culture University Press.]