



## Study on the Order of Vocabulary Output of International Students

---

Wang Xi and Wang Zhimin

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

June 1, 2019

# 留学生词汇输出顺序研究<sup>1</sup>

王希 王治敏\*

(北京语言大学, 北京市, 10000)

**摘要:** 在对外汉语教学过程中, 词汇输出是衡量学习者学习效果的一个重要指标。本文通过向高年级留学生发放 100 份调查问卷, 收集到他们认为最应该掌握的 300 个日常词汇。通过对问卷结果进行归纳与统计, 发现留学生的词汇输出遵循场景线索、词类范畴线索以及词性线索, 同时, 留学生词汇输出也具有一定的规律和特征。其中包括输出的词汇具有可表象性, 双音节词占优势; 词汇输出有初步的语素意识, 体现场景理念, 具有性别差异, 遵循习得顺序。

**关键词:** 基础词汇; 词汇输出顺序; 心理词表; 对外汉语教学

## Study on the Order of Vocabulary Output of International Students

WANG Xi, WANG Zhimin\*

(Beijing Language and Culture University, Beijing, 10000, China)

**Abstract:** In the process of teaching Chinese as a foreign language, vocabulary output is an important standard to measure the learning effect of learners. This paper collects 100 questionnaires from senior international students and collects 300 daily vocabulary that they think should be mastered. Through the induction and statistics of the questionnaire results, it is found that the vocabulary output of international students follows scene clues, word category clues and part-of-speech clues. At the same time, the vocabulary output of international students also has certain rules and characteristics. The output words have imageability and the two-syllable words are dominant. Vocabulary output has a preliminary sense of morpheme, embodies scene concept, has gender difference, and follows the acquisition order.

**Key words:** Basic vocabulary; Vocabulary output order; Psychological wordlist; Teaching Chinese as a foreign language

### 1 引言

汉语学习者的词汇输出是衡量学习者效果的重要指标。在前人的研究成果中, 对汉语词汇输出的研究成果并不多见, 主要集中于英语的词汇输出。鲍贵(2017)采用实验设计检验输出任务, 主要运用组句和填空, 对英语学习者即时词汇习得的影响, 发现输出任务独立作用于词汇习得。黄燕(2017)调查输出任务类型对相关输入中二语词汇注意及附带习得的作用, 通过不同的输出任务, 例如写作、重构及复述强化等, 发现所有的输出任务都能有效提升学习者对目标词的注意, 从而促进其接受性词汇习得; 同时, 时间对任务效应有调节作用。邹一杰(2011)通过设计词表实验, 发现认知风格和性别都能对记忆产生影响。张晓苏(2012)提出, 成年人的二语心理词表的建立往往要依附于母语的心理词典, 如果能够学习

<sup>1</sup>本研究得到国家社科基金重大项目(18ZDA295), 教育部人文社科基金(16YJA740036), 中央高校基本科研业务费(18YBT03)的资助。

\*本文通讯作者

更多的二语词汇，才有可能建立起单独的二语心理词典。王周炎（2004）提出，语素教学是对外汉语词汇教学的基础，先教一些构词能力强的单音语素，多辨析发音相同的词、近义词时应突出语素教学。

前人学者关于词汇输出的成果，为我们的研究提供了启示，但是二语心理词表如何获取？心理词表输出有何特点？是否有一定的规律？都是需要深入探索的问题。

因此，本文先发放调查问卷，探索心理词表的获取方法，接着对调查结果进行分类归纳与统计。为了获得较为客观、统一的研究结果，先对词汇输出线索进行界定，再对输出线索进行分类，计算其出现频次、最大长度、平均长度等数值，从而进一步探求留学生词汇输出的规律和特点，得出结论。

## 2 基础词汇的调查

本次调查问卷的对象是在北京语言大学就读的来华留学本科生，我们收集他们在学习、生活过程中认为最重要的词汇，并将收集的结果进行分类、归纳与统计。通过观察留学生的词语输出的顺序、频率等相关指标，来考量他们的心理词表，从而对对外汉语教学提供一些资料参考。问卷的具体内容如下：

亲爱的各位同学：

您好！以你学习汉语的经验，你认为留学生最需要掌握的 300 常用词是什么？比如学习、生存、日常交际，日常活动等方面，请把自己的想法填写在表格中，感谢您的合作！

与传统的调查问卷不同，这份问卷是开放式的，主观性强，没有任何提示，目的是让留学生不受任何选项的干扰，自主输出在自然状态下第一反应所想到的词语。这样能够最大限度了解留学生们的认知水平、语言水平，把握他们的真实语言需求。

本次调查问卷在课堂上进行，时间约为 60 分钟。调查对象主要是三、四年级的高年级本科生，共 100 人。之所以选择高年级学生，是因为相较于低年级的学生，高年级学生已经掌握了数量较多的词汇与语法点，表达能力也更强，在输出词汇的时候不会较多地被自身的学习水平所限制。同时，我们对留学生的性别也做了统计。在 100 名被调查者中，男性占总人数的 22%，女性占 78%，可见女性是参加调查问卷的主要对象。在这些留学生之中，大部分来自汉字文化圈，如日本、韩国、新加坡等，少数来自非汉字文化圈，例如美国、挪威、俄罗斯、马来西亚、土耳其等国家。

除此之外，留学生学习汉语时间的长短也是影响词汇输出的一个重要指标，因此我们对留学生学习汉语的时间也做了考察。从调查结果中我们可以看出，学习汉语 2 年至 3.5 年（包含 3.5 年）的学生最多，占总人数的 53%；次之是学习汉语 3.5 年至 5 年的学生，占比 31%；其余学生占总人数的 16%。在这些学生中，学习汉语时间最短的只有 1 年，他们应该是来到中国之后才开始学习汉语；学习汉语时间最长的达到 11 年，我们推测该学生在接受本科教育前就接触了汉语，在本国就开始进行汉语学习。大多数学生学习汉语的时长与年级基本保持一致。

## 3 留学生词汇输出线索考察

在对多份问卷进行观察与分析的过程中，我们发现，许多留学生的词汇输出是遵循一

定顺序的，我们以留学生输出词汇序列为研究重点，对多份调查问卷进行标记，尽可能探寻其中的规律，了解他们的学习情况与心理状态。在留学生输出的 300 个词汇中，我们发现一个较为普遍的问题，就是先输出的词语往往错误比较少，后面出现的词语错误比较多。这种现象具有普遍性，在汉语水平较低或较高的学生中均有体现。为了尽可能避免留学生词汇输出的偏误以及思考时间过长而产生的消极情绪的影响，我们只截取每份调查问卷的前 100 个词语，这部分词语错别字、生造词较少，表达的正确率较高，且能较完整体现留学生词汇输出的思路。

在对留学生词汇输出线索进行探究之前，我们有必要先对留学生的词汇输出线索进行界定。我们建立了一套自己的标注体系，划分了一套自己的词语类别，分别对 100 份语料的前 100 个词进行标注分类。在每一份样本中，凡是有 3 个及以上的词语呈现段落化，能连续出现，且这几个词语有一定的共同点，例如，它们都属于同一词类范畴，或者能在同一场景中出现，或者包含相同语素，我们就把它们看作一个连续的“段落”，也就是一条线索，词语之间用破折号表示连续。例如，在某份语料中出现“服务员——鸡腿——素食——账单”这个输出顺序，我们就认为，这些词语都属于“就餐”场景，这个输出序列就是一条关于“就餐”的线索。值得注意的是，一些词语如果要形成一个序列，表示一条线索，必须是连续的。例如，在 100 个词语中，第 1、2 位的词语分别是“爸爸”“妈妈”，而第 10 位的词语才是“哥哥”，那么这样散落的 3 个词语不能划分到表示亲属称谓的同一类别，也不能看作一条线索。

每个留学生的词汇输出有多条线索，在这 100 份调查问卷中，我们根据每种线索出现的频率进行统计，发现线索可以大致进行分类。同时，对于每一份问卷，我们又对每一条线索的长度（即包含词语数量）进行统计，包含最大长度和平均长度（我们规定形成一个序列的最小值为 3，因此最小长度默认为 3，不纳入表格中）分类统计结果如下：

### 3.1 场景线索

我们观察到，一些连续输出的词语可以归纳为一个场景，因此将场景线索进行分类：

表 1 遵循场景线索的词汇输出情况表

场景	输出序列示例	100 份问卷中 出现频次	占比	最大 长度	平均 长度
学校、课堂、 学习	老师——学生——教室——课 ——课本——学习 课文——考试——发音	88	88%	12	4.45
交通、出行	方向——红绿灯——地图—— 警察 换车——方向——返回——开 车——回程——单程	48	48%	10	4.50
日常交际	你好——谢谢——再见 早安——晚安——你好	33	33%	5	3.52

购物	合适——小码——大码——颜色——商场 退款——超市——卖——买	25	25%	6	3.72
就餐	点菜——结账——饺子——鸡排——排骨 服务员——勺子——筷子——叉子	24	24%	10	4.54
看病就医	医院——医生——得病——感冒 发烧——感冒——打针——药品——调理——休息	11	11%	6	4.10
涉外	留学生——外国人——大使馆 护照——钱包——大使馆——中国——签证	8	8%	5	4.25

从上面的表格，我们可以发现：

(1) 在输出序列中，出现在“学校、课堂、学习”场景的频次最高，输出的长度的最大值也较大，可以看出，留学生的主要生活环境还是学校，对学校这个场景有很强的依赖性，当然，也有可能和留学生完成这份调查问卷的地点是教室有关。

(2) 在上述表格中，我们可以看出留学生最需要的词汇基本包含在人际交往、学习、出行、就餐、就医等几大场景内，尽管每一类中的词有其个性，但依旧有很强的规律。据此，我们可以构建基础词表，并将其运用到汉语的教学与学习中。

(3) 300 个词能够较为全面考察留学生的汉语学习水平以及语言能力。在某个场景里，例如点餐、看病、购物或者交通出行，我们要观察学生能做什么，能否进行正常交际，也就是对“can-do”进行描述。词汇输出的序列在一定程度上能反应这些内容。输出的越全面，从这个场景里概括提取出来的词越详细、具体，说明留学生掌握的就越多，就越能完成这个场景中的交际任务。

### 3.2 词语范畴线索

根据观察与统计，我们发现一些连续输出的词汇可以归纳到一个范畴里，它们都有共同的上位词：

表 2 遵循词语范畴线索的词汇输出情况表

范畴类别	输出示例	100 份问卷中出 现频次	占比	最大长 度	平均长 度
家庭, 亲属称谓	爸爸——父亲——妈妈—— 母亲——父母 哥哥——姐姐——妹妹—— 爷爷	76	76%	7	5.59

个人情感, 状态	快乐——难过——高兴 伤心——舒服——幸福	36	36%	3	3.00
个人装扮	烫发——染发——美甲 皮鞋——卫衣——裙子	7	7%	10	5.71
颜色	红色——蓝色——黑色—— 橘红——白色 棕色——红色——紫色—— 绿色——灰色	5	5%	6	4.20

从上表中, 我们发现, “家庭, 亲属称谓”这一类词语出现的频率较高, 说明人们对社会关系的重视, 对亲密关系、对亲情的依赖。而“妈妈”出现的频率在所有亲属词汇中是最高的, 而且在该序列中的顺序靠前, 可见“妈妈”在人们心中的地位最高。亲属称谓的长度大多是双数, 可见留学生在词汇输出时多采用对举的方法, 这也和亲属称谓类词语成对的特点有关。

亲属称谓其实代表的是一种社会关系。我们在调查问卷中, 发现了较多表示社会关系中人物的词语, 有些能形成序列, 例如前文提到的“老师——同学——学生”, 有些不成序列, 散落分布在留学生的词汇输出词表中, 例如“同屋”“朋友”“恋人”“师傅”等等。据此我们推断, 对于留学生来说, 社会关系也是他们在留学生活中的重要内容。

### 3.3 词性线索

这一类别是从语言本体的角度划分 (为了便于统计, 我们把“包含相同语素”这一条线索也归入此类, 因为均与语言本体有关)。我们发现, 在收集到的调查问卷中, 有一部分是按照词性线索输出词汇:

表3 遵循词性线索的词汇输出情况表

词性	输出示例	100份问卷中 出现频次	占比	最大 长度	平均 长度
时间名词	昨天——今天——明天——后 年 上午——下午——晚上	78	78%	14	4.65
代词	我——你——他——他们 这——那——那里	75	75%	6	3.20
方位, 方向	东方——西方——北方——南 方 上——下——左——右——前 面——旁边——里面	43	43%	8	5.16
包含相同语 素	表示——表演——表扬 电视——电影——电话	24	24%	6	4.00

处所名词	食堂——宿舍——饭馆 公交站——地铁站——清真寺	23	23%	5	3.83
积极意义的 形容词	漂亮——可爱——好看 漂亮——聪明——可爱——勇敢	21	21%	5	3.19
数词	一——二——三——四——五 十——百——千——万	21	21%	20	8.62

在上表中我们可以得到如下发现：

(1) 在所有的调查问卷中，实词多，虚词少。这刚好符合学生学习的特点，也是语言教学的特色。实词里面名词比较多，特别是具象名词，可能因为它们意义简单、实在，容易学习。年级越高，学习程度越好的留学生，输出的抽象名词也越多。我们推测，随着学习的句法结构的复杂程度增加，虚词的学习量、运用量也会增加，因此在输出的词汇中虚词比例也会上升。

(2) 在统计的过程，我们还发现了留学生的词汇输出和他们的文化有关。在收集到的100份调查问卷中，有17份为来自马来西亚的学生所填。在这17份问卷中，有11份问卷都提到了“清真”或者“清真寺”这类词语，而马来西亚国教为伊斯兰教，可见留学生的词汇输出和本国的文化、宗教有密切联系。

### 3.4 词汇输出线索分布数量情况考察

在上面的几个表格中，我们了解到了留学生词汇输出的线索。对于每一位留学生来说，他们是有选择地输出心中的词汇。因此对于每一份调查问卷，一定会有多种输出的线索。我们对100份问卷进行统计，将每份问卷输出的线索数量、该线索数量出现的份数以及占100份问卷的比例整理如下表：

表4 每份问卷输出线索情况统计表

每份问卷输出线索数量	份数	占比
3	6	6%
4	17	17%
5	20	20%
6	13	13%
7	19	19%
8	4	4%
9	6	6%
10	5	5%
11	7	7%

12	3	3%
----	---	----

从上表我们可以看出，每份问卷出现的线索数量最少有 3 条，最多有 12 条；出现 5 条线索的问卷最多，占到 20%，包含 7 条线索的问卷仅次之，占到 19%。据此，我们计算出每份问卷所包含线索的平均数为 6.46 条。可以推知，留学生在填写问卷进行词汇输出时，基本会用到 6 至 7 条完整的线索。

#### 4 留学生词汇输出特点分析

通过上文对调查问卷进行统计性描述，对留学生词汇输出的线索进行探求并梳理归纳，我们发现，留学生的词汇输出存在一些规律和特点。具体内容如下：

##### (1) 输出的词汇具有可表象性

可表象性是指一个词能够引发出对应的心理表象的难易和快慢程度。通俗来讲，词汇的可表象性是指当人们看到一个词的时候，它的形象（包括视觉、声音等）是否容易在人的脑海中出现。在汉语词汇中，具体名词以及专有名词的可表象性较强。反推之，在调查问卷中具体名词出现的频率较高和留学生在填写问卷时先看到的都是具体形象有一定关联，如“桌子”“椅子”“学生”“老师”“钥匙”“电脑”等等，这些都是最先呈现在学生眼前的具体形象。同类例子还有“课本”“眼镜”“自行车”“图书馆”等等。

##### (2) 双音节词在输出词语中占优势

在对调查问卷结果进行统计的过程中，我们发现双音节词语占绝大多数，而单音节词语与三音节及以上词语占比很小。这与汉语的音节特点有关。在现代汉语中，单音语素多，双音节词语占优势。在对外汉语教学过程中，有关词汇部分的教学必然会以双音节词语为主，因此留学生词汇的输入与输出多为双音节词。

##### (3) 词汇输出具有初步的语素意识

在留学生的认知以及学习过程中，对语素究竟有没有一个明确的认知，对语素和词的划分界限是否清晰明确，以及明确到什么程度，在现有的资料和语料基础上，仍是值得探讨的问题。但是在调查问卷中可以大致看出，留学生对语素和词的界限并不明确。除去大部分双音节词和小部分三音节及以上词语外，在剩余的单音节词中，其实包括了一些单音语素，例如“师”“阅”“安”“荐”等。

学习时间较长，汉语水平较高的学生在输出词汇时会按照包含同一语素的词进行排列，以他们已有的水平穷尽输出以后又会选择新的语素进行组合然后输出。可见，学习程度较高的留学生已经有了初步的语素意识，在他们认知范围内，已经可以把语素提取出来作为一个语块和其他语素进行组合。但是在汉语词汇里，复合词占多数，由同一语素组成的词的实际意义差别有时大有时小。

例如在“假山——假发”中，尽管它们都包含语素“假”，可是语素“山”和“发”的意义联系较小，不在一个范畴里。但是“假”的语素义是相同的，都表示“人工的，非天然的”。在“国粹——国土”这个输出序列中，都包含“国”这个语素，表示“国家”，但是这两个词的实际意义联系却不大。“国粹”的含义为“我国文化艺术中的精华”，而“国土”指

“国家的领土”。留学生把这几个词放在一起，可以看出尽管这两个词意义差别很大，但是他们能分辨出语素义是相同的。

#### (4) 词汇输出体现场景理念

例如，在“发烧——感冒——打针——药品——调理——休息”这个输出序列中，可以推断出留学生是按照“看病”这一个场景中可能出现的词汇进行输出的。同类还有“点菜——微信——支付宝——肯德基——鸡排——排骨”是按照餐馆就餐这一场景来输出的。“合适——小码——大码——颜色——好看——商场”这一输出序列是按照商场购物的场景输出的。由此可见，部分学习程度较高的留学生已经可以根据不同的场景来选择词汇进行同类的归纳。

#### (5) 词汇输出具有性别差异

留学生的性别差异也是影响他们输出词语的重要特征。性别不同，关注点和兴趣都有差异。在调查问卷中，女性认为的常用词汇里，有关饮食、服饰、美容、情感类词语较多，例如“好吃”“面包”“短裙”“美甲”“爱情”“暗恋”等词语，而男性的词汇输出中出现运动类词语的频率较高，例如“网球”“篮球”等。在对亲属类词汇的处理方面，女性倾向于列出较多的家庭成员，包括“爷爷”“奶奶”“爸爸”“妈妈”“弟弟”“妹妹”等，而且出现的位置比较靠前，而男性对亲属词语输出的顺序比较靠后，且数量较少，不会列举较多的家庭成员，一般以“妈妈”“爸爸”为主。可以看出，女性对家庭成员、家庭关系有更多的关注与依赖。

此外，男性一般以理性思维为主，同时不太注重个人情感的表达，对一些心理活动、情感不是特别敏感；女性则恰恰相反，以感性思维为主，个人感情丰富且细腻，注重感情的表达。这一点性别上的差异在调查问卷中有较鲜明的体现。据统计，61.5%的女性输出了“个人情感”类词语，例如“高兴”“伤心”“难过”“失望”等，而仅约有27.2%的男性输出了“个人情感”类的词语。

#### (6) 词汇输出体现习得顺序

留学生在学习过程中先掌握的词语或者语块的使用频率相对较高。根据人的记忆规律，先掌握的内容会先进入长时记忆中，在输出时，汉语学习者掌握程度较好的、记忆较牢固的汉语词语往往会被先提取出来。留学生词汇先输出的那一部分其实是掌握程度最高的。在对外汉语教学的教材中，知识点的编排顺序和交际需要有一定关联。在交际过程中，使用频率较高的交际用语以及日常沟通语言会先出现，因此，这些内容涉及到的词语也会在部分教材中先出现，就会被学生先习得，也就会被学生先输出。例如，在录入语料的过程中，发现“你好”“谢谢”“再见”“对不起（短语误为词，在这里，类似的语块全部当作词来统计）”等输出频率较高，因为它们是最常用的交际用语，在学习时最先掌握，最先进入到长时记忆中。

## 5 结语

本文向100位留学生发放问卷，调查他们的汉语词汇输出情况。通过对他们词汇输出结果进行统计归纳与分析，我们发现，留学生的词汇输出遵循一定线索，例如场景线索、词类范畴线索以及词性线索等。同时，留学生的词汇输出也体现一定的规律和特点。其中包括输出的词汇有可表象性，双音节词在输出词语中占优势，词汇输出具有初步的语素意识，能

够输出体现场景理念,具有性别差异以及体现习得顺序。我们的研究希望能对对外汉语课堂教学、教材编写以及汉语学习者基础词表的整合提供一些参考。

## 参考文献

- [1] 鲍贵, 李景怡. 输出任务和注释对英语学习者词汇习得的影响研究[J]. 解放军外国语学院学报, 2017, 40(06).
- [2] 鲍贵. 输出任务类型和词语意象性对英语学习者词汇习得的影响[J]. 外语与外语教学, 2016(06).
- [3] 黄燕, 胡新, 王海啸. 任务类型对输出启动下二语词汇附带习得的影响[J]. 现代外语, 2017, 40(05).
- [4] 郜一杰. 言语一表象认知风格对 DRM 范式中错误记忆的影响[D]. 延边大学, 2011.
- [5] 连秀萍. 不同语言输出方式对接受性和产出性词汇习得影响的实证研究[J]. 西南农业大学学报(社会科学版), 2012, 10(01).
- [6] 史有为. 对外汉语教学最低量基础词汇试探[J]. 语言教学与研究, 2008(01).
- [7] 宋佳. 书面输出对外语词汇习得影响的个案研究[D]. 东北师范大学, 2010.
- [8] 王新朋, 孔文, 王永祥. 输出驱动对多维词汇知识习得的影响研究[J]. 外语与外语教学, 2017(01).
- [9] 王周炎, 卿雪华. 语素教学是对外汉语词汇教学的基础[J]. 云南师范大学学报, 2004(05).
- [10] 张晓苏. 认知心理视角下的对外汉语词汇课堂教学研究[D]. 华中师范大学, 2012.
- [11] 朱振方, 刘培玉, 王金龙. 一种基于语义特征的逻辑段落划分方法及应用[J]. 计算机科学, 2009, 36(12).