播客在对外汉语教学中的应用
——以 iMandarinPod 为例

北京语言大学汉语进修学院　Daniel S. Worlton

摘　要　信息时代涌现出了大量前所未有的语言学习工具，其中，播客能满足认知和二语习得理论的基本要求，尤其在对外汉语教学领域。本文分析了语言教学播客的结构，并对 iMandarinPod 的语料库进行了详细的词汇分析。以词的分布度为标准，发掘出了该播客系列的内隐大纲，并提出一个难度评测算法。通过词汇习得计算模拟，本文证明播客对学习者的词汇增长具有重要意义。

关键词　播客　词汇分析　分布度　HSK 大纲

一、引言

随着汉语热的升温，高校、孔子学院、培训机构及私人辅导的汉语教育市场日益旺盛。与此同时，各种各样的网站、软件、博客、在线电视及广播等信息时代的网络资源给学习者提供了丰富多样的汉语学习资源，也使学习者的学习方法和途径更加多元化。那么，这些资源能否使汉语学习者在脱离传统课堂和汉语语言环境的情况下仍然成功地掌握汉语？本文在简单回顾第二语言习得相关理论的基础上，讨论汉语学习的特点，并介绍汉语教学播客在汉语习得中的大量与可理解性输入的作用。

二、从认知角度看二语习得

应用语言学界普遍认为在一语和二语的习得过程中绝大部分的词汇是在日常的听读活动中无意识地掌握的（Hulstijn，2003）。这样的无意识学习又称为内隐学习，又称为伴随性学习。就词汇学习而言，伴随性假说指学习者在不同语境中多次接触某个词以后，会自然而然熟悉这个词（Coady，1997）。美国多项关于学生词汇量的研究发现学生的词汇并不仅仅来自课堂教学，还有相当数量的词汇是

伴随性习得的，因此，伴随性学习已经为母语词汇习得领域公认（Nagy 等，1985）。研究者经常提到的 Krashen（1989）输入假说就是伴随性学习的特例。Krashen 认为语言习得是无意识的过程，并且强调只有这种情况才叫“习得”。输入假说对输入提出两个要求：一是要有大量的语言输入，二是输入必须是可理解的。为了满足大量输入的要求，要保证阅读材料的丰富性才能吸引学习者的兴趣，并鼓励他们大量阅读。关于可理解性，Laufer（1989）提出学习者必须掌握文本中 95% 的词才能理解其意义，如果能达到 97% ~98%，读者就能轻松理解全文（Hirsch & Nation，1992）。换句话说，只要能理解 95% 以上的词，学习者就能够通过语境来猜测生词承载的主要信息。因此，减少生词带来的压力非常重要。在英语中（含文学作品），要达到 95% ~98% 的词汇覆盖率需要掌握 3000 ~5000 词族（word family），或者 5000 ~8000 词项（lexical item）（Nation，1990）。对于初、中水平的学习者来说，大量输入和可理解性输入存在一定的矛盾，因为词汇要通过广泛阅读而习得，但是学习者需要先掌握不少词才能理解更多篇的文本（Coady 1997）。

大量输入不仅对词汇习得起重要的作用，而且是信息加工理论重点关注的问题。在理解之前，阅读（或听）的过程伴随着一系列的认知加工，如分词、识别词、提取语义、进行语法分析等。这些认知模块随着训练的增加，进行重组和自动化（Segalowitz，2003；McLaughlin & Heredia，1996）。自动性认知加工使阅读者（或听者）的理解又快又准。因此，二语学习者只有通过大量的输入增强训练，才能进一步提高语言水平，并且攻克越来越难的输入。

三、汉语学习的特点

世界上绝大部分的语言使用表音文字，而汉语使用表意文字，即汉字。上万个汉字需要一个一个去学习其写法、读音、意义以及构成的词。为了减轻汉语学习者畏难情绪，有一种说法认为只需要掌握 2000 ~3000 汉字就能读懂一份报纸。“常用字表”包含的 2500 字在普通文本中的覆盖率达到 97.97%，再加上另外 1000 个“次常用字”的 1.51% 覆盖率，一共达到 99.48% 的覆盖率（邢红兵，2007）。那么，理论上讲，掌握 3000 左右最高频汉字就可以看懂大部分的阅读材料，这比掌握英语的 3000 词族更有效。但实际上，这个说法是有问题的。词才是语言中的最小的独立承载意义的单位，而不是汉字。2500 常用字能构成 46814 词，加上 1000 次常用字能构成 5.2 万余词（周上之、金朝炜，2014）。这样的词

汇量对母语者来说，已经很不错了。在汉字习得方面，邢红兵（2007）分析了5套小学语文教材，发现每套教材平均含3245字种。可见，汉语母语者需要6年专门的训练才能掌握3000多汉字。对于初学者来说，在进入伴随性学习的阶段之前，这2500个字及其构成的几万词汇量是非常高的门槛。

汉字音和形的分离对母语为拼音文字的学习者带来了更多的困难。Brown 和 Haynes（1985）研究过来自日语、西班牙语及阿拉伯语背景的以英语作为第二语言的学习者。以表音文字（即西语和阿语）为母语的学生阅读能力和听力的相关性很高，日语母语者的阅读能力和听力则几乎没有相关性。这说明，如果母语和二语都使用拼音文字，那么其二语的听读技能之间有传递效应。然而，对于拼音文字母语者来说，汉语的两个模式（听说与读写）是相对独立的系统，因此，广泛阅读对他们的听力不会有明显的影响。

另外，音节作为汉语语音系统的基本单位，本身的一些特点对听力有干扰。平均每个音节有16个相似邻居（李梅秀等，2014）。邻居指通过增加、替换或者删减一个音节中的任意一个音位得到的合法音节。由于不同语言音位的划分不同，再加上汉语音节的邻域密度普遍很高，所以二语者在听辩汉语语音时有很大的困难。除此之外，汉语存在大量的同音字，一个音节最多能对应92个汉字，平均对应8个汉字（苏新春、林进展，2006）。而且，每个汉字还对应多个义项和构词数。这就使得二语学习者从语音通达语义的加工过程更加困难。

综上所述，虽然有个别特例，但是我们认为除非二语者已经达到较高的语言水平，否则真实语料和自然环境对他们的伴随性学习作用不大。因此，要解决初学者大量输入与可理解输入之间的矛盾，须利用适合学习者水平的材料为其循序渐进的学习搭建阶梯，比如，分级阅读和课本。另外，学习者还需要一个训练听力的工具。

四、播客

播客是指由个人或者机构录制的音频或视频，它们通过网络的新媒体方式传播。中国学者很早就意识到播客对教学的重要意义（孙伟彦，2008；赵子剑等，2008）。白乐桑、栾妮（2012）具体论述了播客在对外汉语教学领域的突破：播客作为教学辅助工具，改变了“以教师为中心”的教学模式，培养了学生的自主学习能力，并且使培养的重点从传统的听、说、读、写四个技能的培养转变为交际能力的培养。为学生提供了无负担、趣味性的学习途径。播客本身能够保证

足量的语言输入，增加口语表达的机会，增加学生听、说汉语的时间。这些文章提到了一些播客的优点，实际上，白乐桑和栾妮（2012）并没有完全理解播客的性质，他们将播客理解为一个用来辅助课堂教学的学习工具，但播客其实是一个可超越课堂的学习途径。

4.1 汉语教学播客

本文不详细讨论如何建立一个完整的播客分类法，只讨论语言教学播客。我们将语言教学播客粗略地分三大类：一是短语练习，即把目的语和教学用语（通常为听者的母语）的句子对应，并让播客的收听者自己重复目的语的句子。这种基于听说法的播客教学类似旅游者常常买的常用语手册，每节播客的内容围绕某个生活中的话题展开。如，OneMinute Mandarin 和 Chinese Survival Phrases。二是从易到难、循序渐进的系统性教程。随着课程的进展，播音员（即老师）把越来越多目的语融入教学用语，直到相对高级的阶段，完全用目的语来解释新的内容。如，Chinese Learn Online（基于台湾），第一集没有汉语，最后一集没有英语。三是分模块教学。即把每一节的内容作为一个独立教学模块。如，iMandarinPod（基于天津），Chinese Pod（基于上海）及 Popup Chinese（基于北京）。前两种播客只不过是把已有磁带语言教程（如 Berlitz 或者 Pimsleur 公司的 Language Learning Tapes）改成新媒体传播方式。我们感兴趣的是第三种，下文“播客”只指第三类语言教学播客。

4.2 播客的结构

大部分的语言教学播客有一个共同的结构。有两个人兼主持人和老师的身份，其中一个是目的语母语者，另一个是高级二语者。他们首先播放一个场景对话。重复播放几遍之后，主持人用教学用语逐句解释对话内容，对生词、语法点及文化点进行说明解释，并提供一些例句。播客结束之前，会再次播放对话以巩固所学内容。可见，播客基本包含了传统课本里的所有元素。会员可以上官方网站做练习或者获得更多的增值服务。因为播客音频一般用 MP3 文件，该文件格式可带文本内容，可以在 MP3 或者手机播放时边听边看。

4.3 播客与传统教材的对比

播客与传统教材（即课本）的最大区别在于播客的单元没有固定的大纲和严格的顺序。不过，播客缺乏大纲并不是劣势，而是优势。因为没有大纲拘束

缚，播客制作者可以追踪时尚，讨论时事和热点话题。比如，2008 年北京奥运会期间，Chinese Pod 当时制作了一系列关于运动的播客，既体现了热点话题，又具有时效性。与此相比，课本从编撰到出版需要很长的时间，有时候，课本还没印刷，内容已经过时。另外，作为一种商业模式，播客必须把学生视为客户，重视他们的满意度，才能很好地应对市场竞争。因此，虽然播客制作者没有跟一个固定的大纲，但至少会有一个内隐的大纲，即对理想客户，什么样的内容既实用又有意思。另外，播客制作者还会给播客分级以供学习者选择适合他们的难度和兴趣的内容。因为制作者不知道收听者的身份，有可能是大学生、商人等等，因此必须设计各种不同难度和题材的话题，甚至包括一些带争议性的内容，以保证不同身份的收听者都可以找到满意的学习材料。再加上汉语教学播客中有一个主持人是汉语学习的亲历者，比较了解一般学习者可能遇到的问题，从学习者的角度来设计播客的内容，有助于学习者顺利地克服学习难点。

一家播客机构运营几年之后，会积累成百上千的播客集，远远超过最大规模的课本系列。以《新使用汉语课本》为例，一共有 6 册，合计 72 节课，每节课都包含课文、附加阅读、对话及听力练习，合起来还是不到 300 个语段。播客的课文设计互相独立，数量庞大。重复的内容是常用的或高频的，学习者在学习过程中会自然得到强化，而且播客囊括的内容远比课本内容丰富。

五、播客词汇分析——以 iMandarinPod 为例

5.1 iMandarinPod 介绍

为了更详细地探讨汉语教学播客，我们选择其中一个例子进行词汇分析。iMandarinPod（“爱汉语播客”）由天津的一家公司营运。该播客可以分两个系列：慢速新闻广播和课文。课文的话题一般为日常生活对话、俗语、历史故事、现代中国及中华民族传统文化等。与前文介绍的略有不同，“爱汉语播客”没有难度分类，都相当于中级，最适合已有一定基础的学习者。该播客的另一个特点是主持人均为中国人，教学用语全是汉语，即生词、语法点等都以相对简单的汉语讲解。老师会在播放课文之前提两三个问题来引导收听者。由于该播客侧重中国文化且使用僵硬的课文设计方式（“cookie cutter” approach），其内容缺乏灵活性和趣味性。

5.2 iMandarinPod 定量分析

我们通过 iTunes 的播客订阅功能收集了 iMandarinPod 自 2010 年 6 月到 2014 年 7 月播放的大部分播客，并把课文和新闻分别建成语料库。本文只考虑课文类播客。由于该播客的 RSS 只保留最近上传的几十集，在收集过程中因没能及时同步而错过部分课文。MP3 显示的“歌词”就是每节课文的核心内容。我们把这些内容都转换成纯文本文件，进行基本的文本处理并建成语料库。在文本处理过程中，将播客的标题与内容分别导入 Microsoft Visual Foxpro，使用自制的分词软件对课文内容进行分词，并进行局部的分词校对。在统计中不考虑英文字母、数字及符号。建成的语料库一共含 573 节课、101468 词次、165527 字次。

5.2.1 词汇特征

总体上看，该播客语料库的词汇非常丰富，一共有 11711 词种。其中，有 1284 个（11%）超出我们词典的范围①。超出词典的“生词”包括分词错误、新词、专有名词等。最高频的词“的”出现 5047 次，只出现 1 次的词占接近一半的词种，平均频次是 8.66，标准差是 66.52（频次分布见表 1）。

表 1 词汇频次的分布

频次	词种数	频次（10）	词种数	频次（100）	词种数
1	5223	10 – 19	846	100 – 199	74
2	1912	20 – 29	231	200 – 299	26
3	988	30 – 39	103	300 – 399	12
4	649	40 – 49	69	400 – 499	12
5	437	50 – 59	35	500 – 599	4
6	341	60 – 69	33	600 –	12
7	289	70 – 79	26		
8	206	80 – 89	17		
9	155	90 – 99	11		

从表 1 可以看出，高频词（频次大于 100）只有 140 个，中频词（频次为 10 ~99）有 1371 个，低频词（频次低于 10）有 10200 个。另外，播客词汇频次与我们词典内的频率信息有非常高的相关性，r = 0.829。这意味着，播客词汇的分

① 我们参考的电子词典是《现代汉语词典》、汉英词典“CE – DICT”、北京大学语料库制成的词表和北京语言大学语料库制成的词表的合并，包含多于 12 万词条。

布与真实语料有一定的一致性。词汇分布不一致的部分源于播客语料库词汇本身的特点，即部分词汇出现率远远超出规范语料库的频率，包括主持人的名字（如安娜、小云、小丽、张红等）和一些播客或者教学的用语（如大家、俗语、意思、咱们等），这些词的频次都非常高。

针对对外汉语教学，我们可以利用汉语水平考试（简称 HSK）词汇大纲来评判播客词汇的性质。词汇丰富固然好，但是如果播客词汇偏离 HSK 大纲太多，则有两种可能，要么是该播客的词汇对学生来说实用性太小，要么就是 HSK 大纲编得不够合理。首先，11711 词种比旧 HSK 的 8634 个词和新 HSK 的 4993 个词都多①。两个 HSK 大纲只有 4499 个共有的词，HSK 共同词汇分别占新 HSK 和旧 HSK 大纲词汇的 90% 和 52%。“爱汉语播客”同 HSK 大纲只有一部分共有词汇，播客词汇覆盖 53%（4614/8634）的旧 HSK 大纲词和 60%（3002/4993）的新 HSK 大纲词。如果习得 HSK 词汇是学习者的最终目标，那么通过 iMandarinPod 学汉语似乎不是最有效的学习办法，因为通过 iMandarinPod 学习不但遇不到关键词汇，而且还要学习 7000 超纲词，这个结果看起来不太理想。我们将 HSK 分等级，发现低级词的覆盖比较全面，而且该词的出现率相对较高（见表 2 和表 3）。

表 2　旧 HSK 词汇大纲分级在播客库的覆盖

HSK 等级	甲	乙	丙	丁
HSK 词汇量	999	1968	2119	3548
见于播客库	932	1486	1069	1127
百分比	93.3%	75.5%	50.5%	31.8%
播客平均频次	59.37	7.95	3.91	3.00

表 3　新 HSK 词汇大纲分级在播客库的覆盖

HSK 等级	1	2	3	4	5	6
HSK 词汇量	150	147	298	598	1300	2500
见于播客库	148	141	278	538	967	930
百分比	98.7%	95.9%	92.3%	90.0%	74.4%	37.2%
播客平均频次	191.0	75.2	28.4	13.4	5.5	2.93

但是，低级词汇的高度重叠只是说明常用词的定义比较明确，而高级词汇因为是使用频率相对较低的词，所以在真实语料中比较难遇见。

① 各 HSK 大纲报告其含 8822 与 5000 词，但是我们发现各大纲内有重复的词。

因为播客语料库的词都由各集播客分成单元，我们可以使用比出现率更有效的指标——分布度。我们将分布度定义为某个词种出现的播客集数。因为播客语料库只包含 573 集播客，理论上讲，一个词的分布度限定于 1 到 573。与频次一样，播客语料库中分布度最高的词仍然是“的”，分布度为 560，不过，整个播客库里仍然有 13 集不包含“的”①。其平均分布度（标准差）是 5.31（20.22）。把词按照其分布度进行排序，发现我们的分布与频次分布相似（见表 4）。

表 4　分布度

分布度	词种数	分布度（10）	词种数	分布度（100）	词种数
1	6374	10－19	522	100－199	66
2	1870	20－29	147	200－299	19
3	870	30－39	81	300－399	4
4	543	40－49	52	400－499	4
5	313	50－59	36	500－599	2
6	272	60－69	25		
7	189	70－79	19		
8	152	80－89	11		
9	124	90－99	16		

可见，高分布词（分布度大于 100）只有 95 个，中分布词（分布度为 10～99）有 909，低分布词（分布度低于 10）有 10707。分布度与频次有一定的相关性，某个词的频次不可能少于其分布度。因此，出现 1 次的 5223 个词必定包含在分布度为 1 的词的范围内。不过分布度为 1 的词也可能有相对高的频次，如夜市、献血、垃圾、牌坊、空竹、婚帖、邮票等，分布度都是 1，而频次在 15 以上，这些词都跟其所在的播客话题有关。我们认为频次基于分布度的比例可以作为发现主题词的标准。

我们假设分布度代表播客的内隐大纲，一个词的分布度越大，该词越应该优先学习。换句话说，常用的词不是教学目的而是播客制作者心目中学习者的默认词汇。于是，我们可以假设所有分布度大于或等于 10 的词（即 1004 个词）是核心词汇，反过来，分布度越低该词越可能是专用词或跟某个话题有关。为了更详细地理解所谓的低分布词，我们把词按分布度由大到小排列，将词汇范围扩展到分布度更低的词，分别计算累计的词种和词次在整个播客库的覆盖率（见表 5）。

① 但是要注意的是“的”构成其他的词不计算在内。“的”字出现在 569 个播客中。

表5　分布度的长尾巴

最低分布度	累计词种数	累计词种比例	累计词次比例
10	1004	8.57%	69.61%
9	1128	9.63%	71.10%
8	1280	10.93%	72.66%
7	1469	12.54%	74.45%
6	1741	14.87%	76.73%
5	2054	17.54%	78.84%
4	2597	22.18%	81.92%
3	3467	29.60%	85.67%
2	5337	45.57%	91.04%
1	11711	100.00%	100.00%

表5表示如果某个学习者已经掌握了分布度最高的词种，那么他已有的词汇能覆盖整个语料库的78.84%。他需要再学4.64%的词种使覆盖率增加3.08%。但表5的结果仍然不太理想，因为按照伴随性学习（即可理解性）的要求，光按分布度排列词汇学习永远达不到一个95%以上的覆盖率。

5.2.2　播客分集统计

我们以播客的课文为单元进行更多的统计分析。表6列出了播客的长度。

表6　播客长度

	词数	字数
最少	35	55
平均（标准差）	177.08（52.19）	288.78（82.77）
最多	382	576

可见，一节课文并不长，有几十个词到几百个词。这些统计只不过是播客里核心内容的（如对话、故事等）一部分，顶多是整个播客的十分之一。要想进行词汇覆盖统计，最好从独立的播客入手。在表7中，我们将播客词汇分为10级，将词汇从分布度大于等于10的范围逐步扩展到播客库的所有词，以各级词汇为标准分析了573集播客的词汇覆盖率。

表 7　以播客为单位的平均覆盖率

最低 分布度	最低 覆盖率	平均 覆盖率（标准差）	最高 覆盖率	覆盖大于 90% 的播客数
10	39.29	68.9（11.66）	92.44	4
9	40.48	70.41（11.37）	93.28	7
8	42.86	71.98（11.17）	94.29	14
7	44.02	73.80（10.79）	95.45	25
6	49.65	76.12（10.19）	96.10	45
5	51.06	78.29（9.56）	96.43	76
4	53.90	81.42（8.63）	97.73	128
3	59.57	85.25（7.57）	99.32	218
2	64.88	90.75（5.64）	100	363
1	100	100（0）	100	573

从表 7 我们可以看出，在各级词汇标准中，都有“可理解”的播客，不过这里的可理解性标准为 90% 以上的覆盖，而不是 Laufer（1989）的 95%。可见，虽然该播客库算得上适合中级水平的学习者，但是各集播客的难度不同。于是，我们使用分布度对播客库的 573 集播客进行相对难度的分析。如果一集播客包含的词语都是分布度高的词，我们认为其难度不高。同样，为了避免在给某集播课定难度时受到重复率高的主题词的影响，尤其是那些重复率非常高的低分布词，我们设计了以下的函数来计算难度：

分布（j）指词 j 的分布度，出现率（i，j）指词 j 在播客 i 的出现次数。这样一来，我们就可以排列所有的播客了。

$$\sum_{\text{播客i包含的词种j}} \frac{1}{\text{分布（j）}^{*}\text{出现率（i，j）}}$$

最简单的播客（难度 = 2.33）如：

A：大家好，我是张路。

B：大家好，我是李珺慧，很高兴又在这个时间和大家见面了。

A：珺慧，我们今天要学习哪个俗语呢？

B：今天我给大家准备的俗语是万事开头难。张路，你来给大家解释一下这个俗语的意思吧。

A：好的。在“万事开头难”这个俗语里面，万就是一万，我们用它来形容很多，万事在这里可以理解成所有事情。开头，就是开始的意思，难，就是困难的意思，万事开头难，就是所有事情在开始的时候都很困难。

B：张路解释得非常清楚。下面，我们来听三段对话，学习一下万事开头难这个俗语的意思和用法。①

最难的播客（难度=53.28）如：

很久以前，汝南县有个人名叫恒景。他和父母妻子一家人守着几片地，安分守己地过日子。谁知天有不测风云，汝河两岸忽然流行起瘟疫，夺走了不少人的性命。恒景小时候曾听大人说过，汝河里住了一个瘟魔，每年都会出来散布瘟疫，危害人间。为了替乡民除害，恒景打听到东南山中住了一个叫费长房的神仙，他就决定前去拜访。

恒景拜访遍了名山大川，终于找到了费长房的住处。他恭恭敬敬地在门口跪了两天，费长房终于收留了恒景，并教他降妖剑术，还送给他一把降妖剑。有一天，恒景正在练剑，费长房走过来对他说：“今年九月九，汝河瘟魔又要出来害人。你赶紧回乡为民除害，我给你茱萸叶子一包，菊花酒一瓶，让你家乡父老登高避祸。”说完，就用手招来一只仙鹤，把恒景带回汝南去了。

恒景回到家乡，在九月初九的早晨，他按费长房的叮嘱把乡亲们领到了附近的一座山上，然后发给每人一片茱萸叶，一盅菊花酒。中午时分，随着几声怪叫，瘟魔冲出汝河，瘟魔刚扑到山下，突然吹来阵阵茱萸奇香和菊花酒气。瘟魔戛然止步，脸色突变，恒景手持降妖剑追下山来，几回合就把瘟魔刺死剑下，从此，九月初九登高避疫的风俗年复一年地传下来。

从这两个例子，可以看出我们对难度的估计有一定的理据。从计算结果来看，这573集播客的平均难度为18.1，难度的分布相对比较均匀，其中332集播客的难度小于平均难度，其他241集播客的难度大于平均难度。

5.2.3　词汇习得计算模拟

为了做进一步的分析，我们对学习播客的过程进行初步的计算模拟。从词汇习得的角度来说，我们觉得核心词汇就是在开始收听iMandarinPod之前学习者至少已经掌握得很牢固的那些词汇。假设一个学习者只知道播客库里1004个高分

① 俗语介绍播客与故事播客的文本相反，虽然我们可能认为三段对话应该是核心内容，偏偏没有转换文字。反而，故事内容里的故事作为核心内容，而没有录入主持人的引言。

布度的词，按照上述难度把播客排列好，一集一集地学习，再假设该学习者的记忆完善，词汇量随着各集播客所含的词的增加而逐步增加。我们可以计算出每集播客的生词比例及学生词汇的增加率。各集播客的生词比例如图 1。

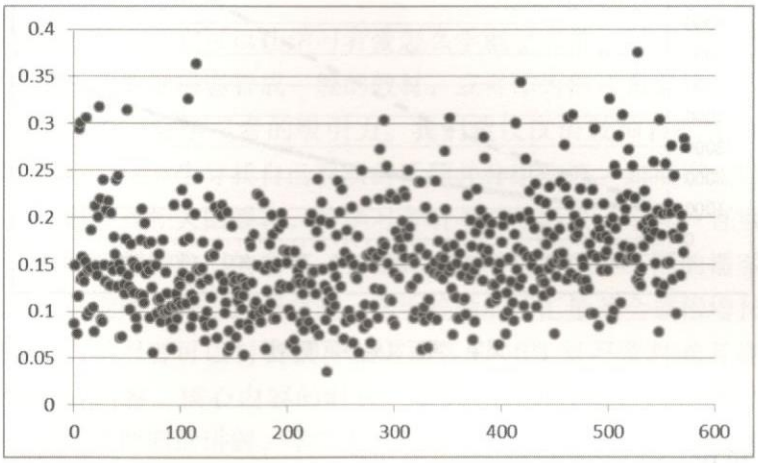

图 1　按难度排列，播客生词比例

图 1 的生词比例就是上文覆盖率的相反概念。根据使用难度进行的排列没有完全消除高生词比例的现象，依然有高达 37.5% 的，也有几十个高于 30% 的，平均比例是 15.58%。观察图 1，在前半段大部分的播客生词比列集中在 0.1 ~ 0.15 的范围内，到中间以后，重量中心逐渐往上挪。计算模拟的前提条件就是把 1004 个词作为熟悉词，那么剩下的 10000 多词种只能出现在 573 集播客中，因此，各集有 18 个生词（10707 词种/573 播客 = 18.686）是正常的。分布低的主题词多次出现就自然而然把生词比例提高了。图 2 显示的是词汇量增加的两条曲线，虚线代表完美记忆，实线表示只在第二次碰见某个词才能记住。

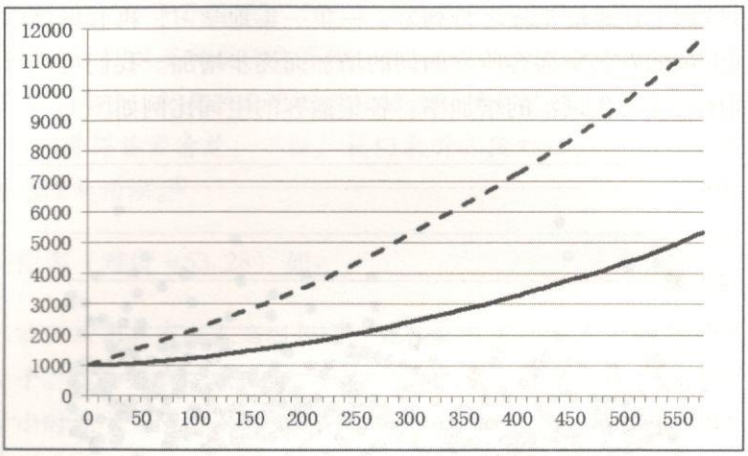

图 2　词汇发展的范围估计

六、讨论

虽然我们从几个不同的角度考虑了词汇要求与覆盖率，结果显示，在 iMandarinPod 播客库里，只有少数例子能达到内隐学习的理想值，即 95%，但是这并不代表语料有问题。其实，作为教学材料，各集播客包含介绍与解释。假设学习者对教学用语的理解没有问题，那么各集播客顶多包含几十个生词，而且这些词的重复次数比我们统计的还多。尽管如此，就词汇而言，播客库在保持一定的真实性的同时，丰富性也很高，所含词汇比 HSK 大纲广。虽然很多词没有理想的重复率，但仍然对学习者的伴随性学习有一定的影响。

如果学习者经常收听该播客的话，应该能够满足 Krashen 大量和可理解性输入这两个条件。可理解性不但是词汇的问题，而且也涉及到学习者的注意力。一般每集播客的内容就几百字的长度，整集播客时长只有 12.89 分钟。一集播客不需要花费太多精力。这也意味着，学习者可以保持足够的语言输入，如果能坚持的话，一个小时内可以听 4 ~ 6 集播客。在 4 年内收集的 573 集播客也够学习者多听多选。其实，学习者不用花很长时间收集材料，一次性从该公司买一套以往的播客就可以了。

播客也是一种可以创造语言环境的工具，学习者可以随时随地带 MP3 反复练习，不用非得来华留学，也不用担心真人说得快或者带有方言口音的问题，这样，学习者一直学习一个相对标准的语言材料。另外，学习者可以参考 MP3 “歌

词”，边听边看，同时训练听力与阅读。

七、结语

播客的好处就在于它可以让学习者爱怎么学就怎么学。因此，要将播客引入课堂需要谨慎。不要把播客看成一般的教材，或者作为作业布置给学生。因为这样做就会使播客失去对学习者的吸引力。我们建议汉语教师自己先熟悉几种播客，推荐给学生，让学生寻找自己喜欢的话题并分组交流。如果正式课程涉及到某个话题，可以寻找相关的播客，给学习者听不同的语段。与其一直重复课本附带的录音，不如听几段类似的播客。如果每个同学自己选了不同的播客，每个学生可以做一个小老师，分组进行交流，学生接触的词汇量就会成倍增长。为了提高学习的效率，学习者可以付费获得相关播客的 MP3 复习资料或者使用音频编辑软件提出核心内容。核心内容的时长一般为一分钟左右，内容总比解释有意思，而且已经听了两遍的讲解，学习者已经知道其意思，这些核心内容就成了可理解性输入了，反复收听会促进学习者语言水平的提高。

参考文献

白乐桑、栾妮（2012）博客教学在对外汉语教学改革中的重大突破，《汉语学习》6：82－90。

李梅秀、邢红兵、舒华（2014）汉语音节结构及领域分布情况统计，《数字化汉语教学》，北京：清华大学出版社，454－465。

马玛丽、姜学军、王者丽（2012）运用播客实现高中英语远程互动学习探索，《理论观察》75：178－179。

苏新春、林进展（2006）普通话音节数及载字量的统计分析——基于《现代汉语词典》注音材料，《中国语文》3：274－288。

孙伟彦（2008）基于播客的教育应用探究，《软件导刊：教育技术》11：88－89。

邢红兵（2007）《现代汉字特征分析与计算研究》，北京：商务印书馆。

赵子剑、王琨、龙吟（2008）应用播客进行教学探析，《保定学院学报》21（2）：119－120。

周上之、金朝炜（2014）基于计算机统计的字——辞教学数量研究，《数字化汉语教学》，北京：清华大学出版社，209－218。

Brown, T, & Haynes, M. (1985) Literacy background and reading development in a second language. In T. H. Carr (Ed.), *The Development of Reading Skills*, San Francisco: Jossey－Bass, 19－34.

Coady, J. (1997) L2 vocabulary acquisition through extensive reading. In J. Coady and T. Huckin

(*Eds.*), *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*, Cambridge: Cambridge University Press, 225 – 237.

Hirsch, D. & Nation, P. (1992) What Vocabulary Size is Needed to Read Unsimplified Texts for Pleasure? *Reading in a Foreign Language*, 8 (2): 689 – 696.

Hulstijn, J. H. (2003) Incidental and Intentional Learning. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition*, Oxford: Blackwell Publishing Ltd, 349 – 381.

Krashen, S. (1989) We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal*, 73 (4), 440 – 464.

Laufer, B. (1988) What percentage of lexis is essential for comprehension? In C. Lauren& M. Nordman (Eds.), *From Humans Thinking to Thinking Machines*, Clevedon, UK: Multilingual Matters, 316 – 323.

Laufer, B. (1997) The lexical plight in second language reading. In J. Coady& T. Huckin (Eds.), *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*, Cambridge: Cambridge University Press, 20 – 34.

Nagy, W. E., Herman, P. A., & Anderson, R. A. (1985) Learning Words from Context. *Reading Research Quarterly*, 20, 233 – 253.

Nation, I. S. P. (1990) *Teaching and Learning Vocabulary*, Rowley, MA: Newbury House.