

计算机自适应性汉语考试的开发研究

——以词汇试题开发为例——

侯仁锋・浅野雅树・丸山浩明

摘要 我们正在开发一个基于网络的计算机自适应性汉语考试(C-CAT: Chinese Computerized Adaptive Test)。该考试是依据项目反应理论设计的自适应性考试,其主要特征是,计算机会根据每个考生的能力自动分配给最符合该考生能力的试题,由此对考生能力做出绝对性评估。考试的开发,首先有必要弄清考试对象,该考试的主要对象定位于日本的汉语学习者。该考试由词汇、语法、听力和读解四部分组成,本文主要报告词汇试题的开发过程,着重探讨了词汇该考什么和怎么考:在开发过程中,在积极利用先行研究成果的同时,根据词的使用频度、亲和度及其话题性等因素选定最基本和重要词汇;发挥计算机优势,开发利用图片、照片或动画形式的多媒体形式试题,使考试更接近真实。

关键词 计算机自适应性考试 项目反应理论 测试对象 词汇测试内容 词汇测试形式

コンピュータ適応型中国語テストの開発について

——語彙問題アイテム開発を例にして——

要旨 中国語能力を、インターネット上でコンピュータによって測る C-CAT (Chinese Computerized Adaptive Test) の開発を進めている。C-CAT は項目応答理論に基づいて構築するアダプティブ(適応型)テストであり、個々の受験者の能力に合致した問題が出題され、結果(能力)は絶対評価で示されることが主な特徴である。テストの開発にあたり、まず受験対象を明らかにする必要があるため、このテストの受験対象は主に日本人中国語学習者であると位置づけた。テストは語彙、文法、聴解、読解の4分野で構成されてい

る。語彙アイテムの開発にあたっては、何をどのように測定するかの方針を決め、既成の研究成果を援用しながら、基本的に重要な語彙を取りこぼさないよう、使用頻度や親密度¹⁾や話題性も考慮したりする選定過程と、コンピュータの特性を活かせるものとして、イラスト、写真等を利用したデジタルアイテムの開発についての研究報告である。

キーワード コンピュータ適応型テスト 項目応答理論 試験対象 語彙テストの内容範囲 語彙テストの形式

1. 概要

计算机自适应考试 (CAT: Computerized Adaptive Test) 是依靠项目反应理论²⁾ (IRT: Item Response Theory) 和计算机技术的支撑而发展起来的新测试形式。这一新测试形式于网络时代愈发显现出了其优点, 因而被应用于越来越多的考试中。

自适应考试的基本原理, 可以追溯到最早的自适应性测试—罗德测验, 在罗德测验中, 呈现给考生的下一个测验项目 (试题、题目, 下同) 是根据他

-
- 1) “親密度”与中文摘要中的“亲和度”同义, 是指一个词与其他词的结合或被结合的程度。
 - 2) 项目反应理论也称潜在特质理论或潜在特质模型, 是针对经典测量理论的局限性 (弱点) 而提出来的, 是一系列心理统计学模型的总称。这些模型的目标是来确定潜在心理特征 (latent trait) 是否可以通过测试项目被反应出来, 以及测试项目和考生之间的互动关系。即, 项目反应理论认为考生在测试项目上的反应和成绩与他们的潜在特质有必然关系。项目反应理论的最大特点是: 找到了一条项目特征曲线 (ICC), 并且以数学函数表达式 (数学模型) 来描述它, 逼近它。不同的数学模型对曲线有不同程度的逼近, 也会有不同的参数。项目特征曲线包含两个方面的参数: (1) 对测试项目的特征进行刻画的项目参数, (2) 对考生特征进行刻画的潜在特征或特质参数。因此, 项目反应理论中测验的项目参数具有跨群体不变性, 具有恒久性的特点。所以所测出的考生成绩具有相对的绝对性, 可在同一个尺度下在时间轴上自己与自己比较, 也可客观地与其他考生进行比较。项目反应理论的这一优良性质为建设大型题库, 编制各种测验提供了方便, 也使计算机自适应性考试的推出成为了可能。但是, 虽然项目反应理论具有很大的优势, 它也有自身的不足: 其中之一就是, 由于受到苛刻的假设限制, 必须要有大样本进行配合, 否则精确性不高。即每一道试题都要至少经过几百人的预测获得参数。这也正是本项目开发的困难所在。

对前一个项目的反应决定的，即对各种不同能力水平的考生给予难度适宜的测试项目。随着计算机技术的发展，后来出现了计算机辅助测试，但这种机辅测试仍以经典测量理论³⁾(CTT: Classical Testing Theory) 为依据，多为纸笔测试形式的电子版，除了测试媒介的改变外，并无实质性进步。直到70年代末80年代初，项目反应理论与计算机技术相结合，出现了真正意义上的计算机自适应测试，现在又基于因特网施测，使测试更加有效、更加高效，也更加客观和准确。

如上所述，所谓自适应性考试，是指计算机根据考生答题情况（对否），推断其能力，提供最适合该考生能力的试题⁴⁾。即考试“量体裁衣”，“因人施测”。这个原理，打个比喻说，如同视力检查。所以，自适应性考试如同检查视力一样，通过交替出难度大（难题）的试题和难度小的试题（易题），来检测出考生能做对的最难试题。即为其最大能力（其水平）。

图 1 是自适应性考试的试题推移与推定能力值变化的示意图。○表示答

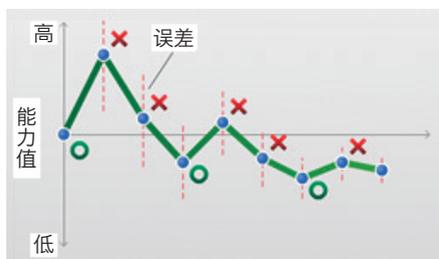


图 1 自适应性考试的试题选择推移与能力值确定的示意图
(参照今井新悟编著『J-CAT オフィシャルガイド』制作)

- 3) 经典测量理论，亦称真分数理论，是一种以考试实得分数为前提条件，以真分数和平行复本为基本概念。在这一理论的基本框架内，建立起测验项目的计量学指标，如测验的信度、效度、题目的难度、区分度等，并以此来筛选试题，编制试卷或建立题库，及其考后项目分析等，为考试走向现代化做出了极大贡献。但在心理和测量领域，人们一般认为经典测量理论有四个弱点。第一是样本依赖。这是经典测量理论的最大弱点，考生特性和测验特性相互依赖：测验或题目的难度依赖考生，而考生的能力水平又依赖于测验或题目的难度。第二是信度和误差问题。第三是无法知道考生在单个题目上的表现。第四是不能“因人施测”。因此被认为测试欠精确度。
- 4) 这种试题已经不是一般意义的试题，而是通过大样本预测后被赋予了参数的试题。

对，×表示答错。答错后会给出一道稍易的题目，如果答对，就会给出一道稍难的题目，这样随着难易度变化逐渐越小，误差也就会越来越小，当收敛到接近无波动处便是最终能力值（水平）。

是为计算机自适应性考试的基本原理。

CAT作为一种新的测试形式，主要有下列优点：(1)是“因人施测”的考试，题目的针对性强，考生需要做答的题目比传统考试少，可提高考试的效度和效率。(2)通过题目调整多次估计考生能力，可提高测量精度，即信度。(3)考试基于项目反应理论设计，可实现不依靠考生群体对考生做出绝对性评估。(4)对考试时间的限制小，可以随时进行考试。(5)可以测试各种水平考生的能力，在一个尺度上或获得到当前的能力水平或定期观察学习成果的发展。(6)能及时得到考试结果，考试结束后即显示（发送、打印）成绩。(7)大量节省印刷、运输等资源，故被谓之绿色考试。

2. 考试对象

无论什么考试，在设计之前必须首先弄清考试对象。因此，考试对象是考试设计的最重要依据之一。考试对象涉及两个方面，一是数量，二是质量（考生水平）。这里首先有必要规定，该考试的主要对象是日本的汉语学习者。但因为考试是放在网上的，也不排除任何国家和地区的考生也可利用该考试。尽管如此，日本的汉语学习者是该考试的基本对象。毫无疑问，这里即便只限于日本的汉语学习者，其水平也有中高低之分及其比例多寡。那么日本的汉语学习者是一种什么情况呢？下面我们从数量和质量两方面进行探讨。

2.1 考生数量

首先，据汉语水平考试（HSK）日本网站（<http://www.hskj.jp/index.html>）的相关信息，说「また、日本国内でも中国語の学習人口は増加の一途を辿っており、200万人を突破したとも言われています」（译文：另外，仅就日本国内而言，据说汉语学习人数也在不断增加，已突破了200万人大关）。这可能是一个时间里（比如近几年）的一个学习人数保有量。由此，再根据我们掌

握的相关信息来看，毫无疑问，应该说日本是世界上汉语学习者最多的几个国家之一。这是我们这个考试成立的基础。

其次，据郭春贵教授的论文，说“全日本大学汉语专业的学生每年不超过1000人，而选修2外汉语的学生，每年近16万。可以说是日本学习汉语人数最多的基地。而大部分学生又都是从零开始，……。”

再者，这是一则新消息，据《日本新华侨报网》（网络版 作者：郭桂玲 发布时间：2015/04/09）报道：“近年来，随着全球化经济形势不断发展，日本学生对赴海外留学关注度高涨。日本文部科学省日前公布的统计数据显示，2012年度赴海外留学的日本学生6万138人，较上一年增加2637人，时隔8年增加。其中，将中国作为留学目的地的日本人数最多，超过2万1000人。”这其中，学习汉语者应该不在少数。还可以估计出，有可能会出不少高水平者。

2.2 考生质量

最后，我们再看一下汉语检定考试（中国語検定試験）的参考人数的情况。我们根据汉语检定协会（日本中国語検定協会）在其网上公布的数据，统计了近5年（2010-2014）的实际参考的考生情况，统计数据如表1所示。

表1 近5年（2010-2014）考生实际情况统计

年度	准4级	4级	3级	2级	准1级	1级	合计
2014	6,698	8,977	10,853	5,513	1,481	310	33,832
2013	8,209	12,358	14,555	7,702	2,136	298	45,258
2012	11,552	18,315	18,384	9,004	2,359	265	59,879
2011	11,399	14,993	15,662	7,607	1,767	258	51,684
2010	11,119	15,005	16,463	8,125	2,280	308	53,300
累计	48,977	69,648	75,917	37,951	10,023	1,439	243,953

为了更能一目了然看出考生的实际情况，我们又做成了图2，图2与其说可以更好地看到考生的数量，不如说更能观察到考生的质量及其分布。

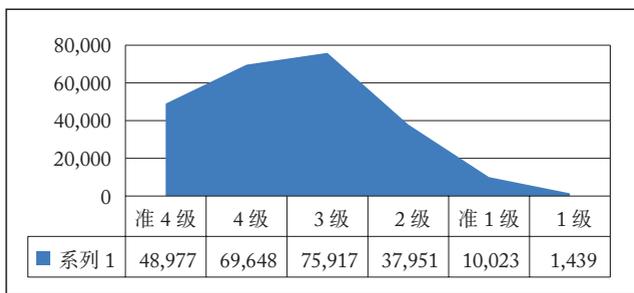


图 2 考生的质量分布 (人)

2.3 小结

综上所述可知三点，一是日本的汉语学习者数量可观，有教学，就有考试的需求，特别是高质量的客观而公正并快捷的测试；二是参加考试者众多，仅汉语检定考试的数据显示，多的年度将近 6 万，少的年度也将近 4 万；三是初、中级学习者居多，占了绝大多数，呈一种质量偏低的基本态势。特别是第三点，将决定我们考试设计的内容范围、范围中区域取题密度，难易度试题量的分布，及其题库中难易度试题所占比例等等。

3. 测试中的词汇能力测试定位

3.1 从结构主义语言学看

本考试系统的开发，参考、借鉴先行研究“日语计算机自适应性考试 (J-CAT: Japanese Computerized Adaptive Test)”的成果，考试系统的测试内容由四大部分组成，即词汇部分、语法部分、听力理解部分和阅读理解部分，从这四个方面测试考生的汉语能力水平。我们认为从这四方面测试是合理的。为什么能这么说呢？

关于对语言的认识，有众多学说，莫衷一是。但至今最有影响力的仍是结构主义语言学说。结构主义语言学说的基本认识是，认为语言是可以分解的，所以把语言从要素和技能上进行了分解，如表 2 所示：

表2 语言技能、要素简表

语言要素	语言技能			
	听	说	读	写
语音	√	√		
表记			√	√
词汇	√	√	√	√
语法	√	√	√	√

(参照王振亚《现代语言测试模型》p. 45制作)

按结构主义语言学观点看，毫无疑问，汉语也可以分解为“语音”“表记”（文字）“词汇”和“语法”四大部分。

从这个表可以看出，“词汇”“语法”这两个语言要素贯穿在各个语言技能之中，和任何一个语言技能都直接相关，通过测试它们可以间接地推测到四个语言技能的能力。另一面，测试“听”“读”，不仅可以测试到通过听觉和视觉获取信息的能力，还可以兼顾测试到“语音”和“表记”能力。从这个表中，我们可以得出这样的结论：一是考试系统从语言要素和语言技能两个方面，设计四大部分全面测试考生的汉语水平，符合语言规律；二是词汇和语法是语言要素的重要部分，应该进行测试。

3.2 从汉语特点看

世界上有5、6千种语言，语言学家从类型学的角度对其进行了大体的分类，汉语被定位于“孤立语”。对此按我们的理解，其最大的依据之一，就是很多时候在语言结构中，无形态变化下词义优先。如下面的例子：

吃米饭。/吃食堂。/吃大碗。/吃父母。/靠山吃山，靠海吃海。

这几个句子语言表层结构完全一样，但不能按照结构做出解释，而必须依据词义才能做出合理解释。显而易见，词义大于结构。所以，词汇在汉语中显得尤为重要。而且这种语言现象在汉语中常见。

再则，如上所述，汉语也可分为其构成的四大要素。对于这四大要素的定位，中国汉语学界也普遍认为，其中词汇最重要。如赵金铭（2005）教授指出：“因为词汇是语言的唯一实体，语法也只能依托词汇才得以存在。”胡明扬

(1990) 教授也有同样的见地：“语言实际上体现了一连串根据语法规则组织起来的词语。所以语音、语法、语义都体现在具体的词语身上。”从测试学的角度，对他们的见地进行解读，我们不妨认为，第一，词汇在汉语中极为重要，应该是测试的重点；第二，对词汇的测试应该是多角度的，应该从读音、写法、词义和用法（搭配）等方面测试。

另外，日本汉语教学重镇东京学艺大学松冈荣志教授根据自己学习的经过和长年的教学经验，在该项目设计、开发探讨时指出：决定“听说读写”四大技能能力的最关键因素就是词汇量。

3.3 词汇能力指什么

综上所述，既然如此，我们则有理由认为，词汇部分是一个测试中的重中之重，如何准确地测试出考生的词汇能力及其词汇量，则是一个考试开发的重要课题。

下面还需要搞清一个问题，对外国人汉语学习者来说，特别对日本汉语学习者来说，词汇能力指什么？或者说包括什么？

从汉语词汇特点考虑，词是由“字”组成，而汉语的字（词）则具有“音形义”三个方面，都是学习者要学习掌握的。由此可见，汉语学习者的词汇能力应该包括如下四点：一是认读汉字能力（音），二是书写能力（形），三是词汇运用能力（词义辨析、搭配运用等），四是词汇量。由这四者构建起其汉语词汇能力。四者之中，我们认为最主要的是词汇量（指能理解和会使用的词汇量之和）。所以考试设计的主要课题是：针对不同水平，测量哪些词最合适？如何测试到考生的词汇量大小？这样就要探讨以下问题。

4. 汉语的最常用词汇是哪些

4.1 词汇计量研究成果

不论是中国的对外汉语教学，还是日本的把汉语作为第二外语教学，都首先要解决的一个基本理论问题是，在基础教学中应该教哪些词汇。对此，先人和时贤做出了不断的探索，取得了丰硕的成果。特别是20世纪80年代以来，

随着网络和计算机技术的发展，汉语词汇计量研究取得了很多大规模更可靠的成果。中日两国等长期的汉语教学实践，与这些词汇计量成果相结合，便产生了若干个教学和测试词汇表。我们通过对一些相关词汇表的考察（侯仁锋·申荷丽「对日本汉语教学初级阶段词表的考察」『日中語彙研究』第4号，2015.3），认为这些词表选词相当科学和可靠，本考试系统完全可以将此作为基本参考词汇表。教什么考什么，当然这些词汇就是测试内容的主要对象。

4.2 主要依据词汇表

经过探讨，本考试系统将主要依据以下几个词汇表划定考试范围、预估难度、选词设计考试。

(1) 教学词汇表

主要依据日本编制的两个：

①高中汉语教学词汇表（718词）

日本高中汉语教育研究会（高等学校中国語教育研究会·高中研）制定的《高中汉语教育指导纲要》（『高校中国語教育のめやす』（1999年版）词汇表，为最基本的常用词及学校用语。

②汉语教育学会（中国語教育学会）制定的词汇表（1000词）

日本汉语教育学会学力基准项目委员会（中国語教育学会学力基準プロジェクト委員会）制定的《汉语初级阶段学习指导纲要》（『中国語初級段階学習指導ガイドライン』）词汇表（2007年3月），是上述词汇表的扩大版，增加了部分常用词及新词，设定为本考试系统初中级水平试题编制的最主要依据词汇表。

(2) 考试词汇表

主要参考新、旧 HSK 两个词汇表：

③新 HSK 词表（5000词）

2009年，由中国国家汉办及孔子学院总部正式推出新汉语水平考试（简称新 HSK），其考试大纲后面，附有各个级别的词汇表，具体如下：

表3 新 HSK 各级词汇量及其水平相当

新 HSK 笔试		国际汉语 能力标准	欧洲语言框架 (CEF)	HSK 各等级对应的课时
级 别	词汇量(个)			
HSK 1 级	150	1 级	A1	按每周 2-3 课时进度学习汉语 一个学期 (半学年)
HSK 2 级	300	2 级	A2	按每周 2-3 课时进度学习汉语 两个学期 (1 学年)
HSK 3 级	600	3 级	B1	按每周 2-3 课时进度学习汉语 三个学期 (1 个半学年)
HSK 4 级	1200	4 级	B2	按每周 2-3 课时进度学习汉语 四个学期 (两学年)
HSK 5 级	2500	5 级	C1	按每周 2-3 课时进度学习汉语 两年以上
HSK 6 级	5000		C2	主要面向掌握 5000 及 5000 以上 常用词汇的考生

注：参照新 HSK 大纲作成。2 级以上为累计词数。表中课时指中国的课时，1 个课时通常为 45-50 分钟。

④旧 HSK 词表 (8822 词)

1999 年推出，分甲乙丙丁四级。实际上新 HSK 词汇表是以这个词汇表为蓝本，精简修订而成的。所以本考试系统主要参考新 HSK 词汇表编制试题。

(3) 汉语词汇使用频率表

⑤现代汉语语料库词语分词类频率表

⑥现代汉语语料库词语类频率表

这两个词汇频率表都是教育部语言文字应用研究所根据大规模语料库而统计出来的，将作为本考试系统编制试题的主要参考词汇表。选词时，以此确认词的使用频率、亲和度及其话题性。

(4) 本考试系统的词汇特色

毫无疑问，不论哪种考试，只要以初中级学习者为主要对象，其测试范围的基本词汇应该是大同小异的，这也是本考试利用以上词汇表的理由。但本考试为了反映时代发展和突出针对性，在选词上具有以下三点特色：一是代表时代发展的网络词汇，如因特网、网络、博客、微信等；二是反映世界一体化的词汇，例如世界上的名城名人小事，纽约、华盛顿、巴黎、伦敦、比尔盖茨、

苹果手机等等；三是反映日本特色的名词，例如富士山、樱花、新干线、相扑、东京、大阪、京都等等。我们将在测试中积极选用这些词汇。

4.3 小结

通过综合考察，这几个词表中越靠前的词，越呈高相关性，就意味着这些词是最常用词，应该是测试的主要对象。尤其是②汉语教育学会词汇表的1000词，与③新 HSK 词表的4级词汇1200词，我们经过对比得知，其一致性（相关）极高，汉语教育学会词汇表的1000词，几乎完全被涵盖在其中。所以在具体选词时，更多的参照新 HSK 词表。同时积极使用反映本考试特色的词汇。另外，为了确保选词（设计的测试点）分布均匀和有代表性，使用词汇频率表从常用性、亲和度、话题性等角度加以确认。

5. 测试题型探讨

5.1 题量及其分布

首先，考试系统设计词汇试题为100题（以100题为一个小型题库，循环使用）。这就意味着如何在5000词中选（组合）出100个最有代表性的词作为测试点。按平均率计算，50词中编制1题，这种划分法，显而易见高级部分将占绝大部分，结果会与考生的情况相乖离。所以有必要考虑水平调整分布，根据低水平考生多的实际情况，越低级别部分的词汇中应该占的比重越大。具体分布如下：

表4 选词大致范围

难易度	易		中		难	
	1级	2级	3级	4级	5级	6级
HSK 级别	1级	2级	3级	4级	5级	6级
词汇数量	150	300 + 150	600 + 300	1200 + 600	2500 + 1300	5000 + 2500
试题数	10	10	20	30	20	10
比例	1/15	1/15	1/15	1/20	1/65	1/250
出题范围	易40			中43		难17

如此划定了选词范围及其题量以后，下一步要做到的就是，如何选出代表性好（包括从常用性、话题性和亲和力等角度考虑）的词作测试点（试题），如何使测试点分布均匀？我们在下节探讨。

5.2 四种题型及难易度分布

本考试系统设计以下四种题型编制试题进行测试。

①看图或单词选读音／看拼音选汉字

②看句子给划线词选图

③选词填空

④看句子给划线词选择适当的解释

（具体试题样题参见后文）

本考试系统（以100题为一个小型题库，循环使用）词汇测试部分按100题设计，按照4个题型平分，每个题型为25题，但考虑到考生水平、日本学习者的特点、初级阶段的汉语特征，以及测试侧重，题量不平均分配，调整为如下分布。

表5 题型分配表

选读音／选汉字	给划线词选图	选词填空	释义选择	合计
30题	20题	30题	20题	100

在此基础上，还需要主要依据考生水平，对难中易不同试题比例做划分。测试理论和实践一般认为，难中易试题分布应该是呈正态分布，即中间大两头小。但如前所述，考虑到考生的实际情况，本考试系统应该取正偏态分布，即中较多，易次之，难较少，呈一种中 \geq 易 \geq 难分布。

表6 难易度大致划分表

选读音／选汉字			给划线词选图			选词填空			释义选择		
30			20			30			20		
易	中	难	易	中	难	易	中	难	易	中	难
12	13	5	8	9	3	12	13	5	8	8	4
40%	43%	17%	40%	45%	15%	40%	43%	17%	40%	40%	20%

整体100题的难中易试题的分布比例是，易题占40%，中等难度试题占43%，难题占17%。在难中易划定后，其内部再分两档，即易1·易2、中1·中2、难1·难2。到这里所讨论的难易度，主要是为了方便命题而划定的。在项目反应理论的设计中，试题经过大范围预测后，难易度最后由预测数据决定，赋予相应参数，所谓的难易度也就消失了。

5.3 四选一形式

计算机适应性考试所有试题设计为四选一形式，并且1题只设计1个考点。四选一试题编制看似容易，实则很难。大量的经验证明，一道科学的四选一试题的产生周期为1年，其过程是：经过培训的命题员命题，专家审题修改，进行预测，修改参数不理想处，再预测，直至各项参数达标（详见：侯仁锋“试析日本汉语检定考试3级试题选项的编制”『中国語教育』第12号，2014）。达标试题的参数为，合格率（难易度）在03-07之间，区分度达0.3以上，干扰项被选率低不能低于5%，高一般不能超过40%，低于5%等于形同虚设，容易造成试题偏易，高于40%，干扰度太强，容易造成四个选项被选失衡，或导致试题偏难。每个干扰项的设计，必须要有依据，不能滥竽充数。

5.4 题型举例及其测试目标

题型1 看图或单词选读音/给句中读音选汉字

这个题型的前者测试目标是，主要测试认读词语的能力，并兼顾测到词汇量。这个题型为了提高考试的信度，在有限的试题中能考到更多的内容，尽管最常用词汇中包括很多单音节词，但考点词语一般选择双音节词。例如：

○看图选读音



- a. diànhuà
- b. tiànhuà
- c. diánhuà
- d. tiánhuà (易)

○看单词选读音

帮助

- a. bānzhù
- b. bānzù
- c. bāngzhù
- d. bāngzù (中)

通过测试能不能掌握一个词语的正确拼音，考察词语认读能力，并通过试题的难易度控制，兼顾测试到词汇量。图画或照片的形式，还有可能兼顾测试到是否掌握了这个词的词义。

这个题型的后者，主要想通过这种间接形式测试到考生的汉字书写能力。其基本想法是，尽管中日语言都使用汉字作为表记，而且也有不少相同的，但毕竟是两种不同语言，在写法上存在差异，不能互相代替，而且汉字本身也比较复杂，也是汉语学习的重要内容之一，所以应该进行测试。我们经过反复探讨和实验，认为汉字写的能力在提供选择的前提下宜从形近、音近、义近、类近、中日差异等角度进行测试。同时这个题型容易拉开难易度，可以兼顾测试到词汇量。例如：

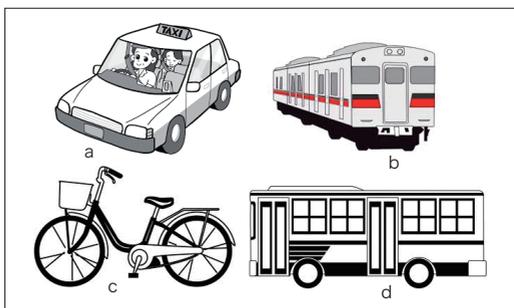
- 他是我们的老 shī
 - a. 师 b. 帅 c. 巾 d. 師 (易)
- 他博 kè 写得非常好，值得一看。
 - a. 课 b. 科 c. 客 d. 克 (中)
- 这场比赛，我们打 yíng 了。
 - a. 胜 b. 赢 c. 输 d. 惨 (难)

题型2 看句子给划线词选图

这个题型通过给划线词选择相应的图片或照片，主要想测试出考生所掌握的词汇量，所以这个题型在试题的难易度上宜加大幅度，拉开距离，使之有明显的差别。即主要通过试题难易度变化明显来测试词汇量。例如：

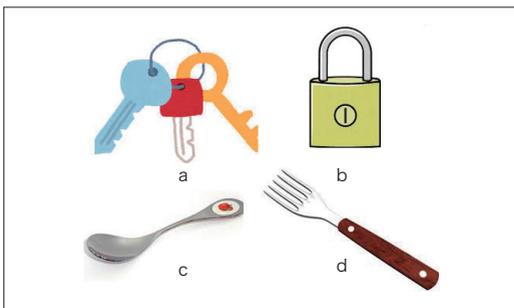
○每天骑自行车去学校。

(易)



○这是谁的钥匙?

(中)



注：视觉图实际均为彩色，或照片或动画。

题型 3 选词填空

这是一个最常见的题型，比较宜于从各种内容和角度进行编写试题，主要测试辨词、选词、用词等词语运用能力，并通过试题的难易度控制，兼顾测试到词汇量。例如：

○有 5 _____ 书。

- a. 本 b. 件 c. 个 d. 张 (易)

○孩子的成绩一直很_____，爸爸妈妈都非常高兴。

- a. 优良 b. 优越 c. 优异 d. 优美 (中)

○在这场全球性的大战中，多少人失去了生命，这是人类的_____。

- a. 悲伤 b. 悲哀 c. 悲痛 d. 悲惨 (难)

题型 4 看句子给划线词选择适当的解释

这个题型主要测试两方面的能力，一是词汇的理解能力，二是词汇量。这个题型比较容易编写出较难的试题，所以可以期待在测试词汇量上发挥更大的

作用。例如：

- 他在学习汉语方面没少花功夫。
 - a. 力气 b. 费用 c. 成本 d. 时间 (易)
- 你这孩子怎么养成了这种坏毛病？
 - a. 习惯 b. 风气 c. 作风 d. 习俗 (中)
- 每个人都带着沉重的心情来到了这里。
 - a. 严重 b. 难过 c. 庄重 d. 深刻 (难)

6. 结语

任何一种考试，在设计时，首先必须明确两点：一是考试目的是什么？二是考试对象如何？在此基础上，考虑考什么和怎么考。所以本文首先简单地提及了基于网络的计算机汉语考试的特点，及其目标，进而探讨和分析了日本的汉语学习者的数量和质量。针对这两点，我们探讨并设定了明确的考试范围及其详细的取题比例，以谋求考试有良好的内容效度。在此基础上，设计了4种题型，从不同的角度，以期全面地测试到考生的词汇能力水平。在试题的制作上，充分发挥计算机的优势，利用彩色图片、照片、动画等数码技术，提高试题与真实语言的接轨程度，推出一些纸质测试无法实现的考试形式（手段），以期进一步提高考试信度。

参考文献

- 今井新悟 (2012) 『J-CAT オフィシャルガイド』 ココ出版, pp.1-11
王佶旻 (2011) 《语言测试概论》北京语言大学出版社, pp.311-315
王振亚 (2009) 《现代语言测试模型》河北大学出版社, pp.112-125
张凯 (2013) 《语言测试概论》商务印书馆, pp.139-162
李珠·姜丽萍 (2008) 《怎样教外国人汉语》北京语言大学出版社, pp.162-173
郭春贵 (2014) “对外汉语教学的突破口” 『広島修大論集』第54卷第2号, pp.37-46
侯仁锋 (2014) “试析日本汉语检定考试3级试题选项的编制” 『中国語教育』第12号, pp.136-150
侯仁锋·申荷丽 (2015) “对日本汉语教学初级阶段词表的考察” 『日中語彙研究』第4号, pp.33-47

计算机自适应性汉语考试的开发研究

付記：本稿は、科学研究費助成金基盤研究(B)（研究課題：コンピュータ適応型中国語テストの開発と検証、研究番号15H03225）による研究成果の一部である。

語彙部門の問題アイテムの開発にあたり、愛知大学中日大辞典編纂所研究員、西安交通大学教授、県立広島女子大学名誉教授であられる顧明耀先生から色々のご助言、例示をいただきました。ここに厚く御礼を申し上げます。

侯仁鋒 Hou Renfeng 県立広島大学教授 専門：中国語教育・言語テスト
E-mail: hourenfeng@gmail.com

浅野雅樹 Asano Masaki 慶應義塾大学准教授 専門：中国語学・語彙教育
丸山浩明 Maruyama Hiroaki 県立広島大学教授 専門：中国語学・中国古典小説