

## 基于“国别化汉语中介语语料库”的字表研制\*

全香兰 胡晓清

鲁东大学国际教育学院, 山东烟台 264025

[xianglanyt@126.com](mailto:xianglanyt@126.com) [xiaoqingtyt@126.com](mailto:xiaoqingtyt@126.com)

**摘要:** 在对外汉语教育领域, 随着国别化研究的逐步深入, 汉字、词汇及语法大纲面临着新的挑战。本文以国别化汉语中介语语料库为基础, 就面向韩国学生的汉字教学用字表的研制进行探讨和研究。文章主要探讨以下四个方面的内容: 字表要收入的汉字的量; 字表的选字原则; 字表的等级划分; 字表的标注内容; 字表在教学中的应用。

**关键词:** 汉字字表 对韩汉字教学 国别化

## Chinese Character List Design Based on a Country-specific Interlanguage Corpus

QUAN Xiang-lan, HU Xiao-qing

College of International Education, Ludong University

**Abstract:** Within the teaching-Chinese-as-a-foreign-language field, the Chinese characters, vocabulary and the grammatical program are facing new challenges with the gradual further research of country-specific teaching. This paper explores and studies the developing of the Chinese characters teaching list to Koreans, with a focus on the teaching to the Korean students. It mainly consists of four aspects: a. The quantity of the Chinese characters the list concludes; b. The principle of selecting the Chinese characters; c. The classification of the characters; d. The content to be tagged of the list and e. Application of the character list in teaching.

**Key Words:** Chinese characters list; Chinese characters teaching to Koreans; country-specific teaching

### 0 引言

关于国别化汉字字表, 学界虽有不同的观点和看法, 我们认为针对韩国学生有必要制定国别化汉字字表。据国别化中介语语料库<sup>①</sup>统计, 韩国留学生使用的汉字与《汉字等级大纲》汉字的重合率分别是甲级字 75.38%, 乙级字 47.26%, 丙级字为 28.29%, 丁级字为 33.57%。除了甲级字, 其他级别的汉字重合率低于 50%。这个数据告诉我们, 针对韩国学生的国别化汉字教学用字表的制定是必要的。《对韩汉字教学用字表》(下面简称《对韩字

\* 本文获国家社科规划项目“多层偏误标注的国别化汉语中介语动态语料库建设与研究”(项目编号: 11BY050) 以及鲁东大学博士人才引进项目“韩国学生汉字习得研究”(WY2013015)的资助。

<sup>①</sup> 该语料库为国家社科规划项目“多层偏误标注的国别化汉语中介语动态语料库建设与研究”的研究成果。

表》)的研制将有助于韩国学生结合本国汉字教育和汉字词基础,更有效地掌握汉语汉字。

## 1 字表的汉字数

现有对外汉语教学用汉字大纲所收录的汉字为 2905 个,其制定主要依据是 3500 个常用汉字。“1988 年,国家语委在前人反复研究和实践的基础上,加以综合分析和科学界定,将常用字分为两个等级:一级常用字 2500 个,二级常用字 1000 个,一、二级常用字共 3500 个。经实践验证,一、二级 3500 个常用字种在二百多万字的检验语料中的覆盖率达到 99.48%,一级常用字 2500 个字种覆盖率达 97.97%。”<sup>②</sup>刘英林、宋绍周(1992)认为根据统计资料,这 2900 个常用字,其累计频率高达 98%以上。张凯(1997)、李清华(1999)、姜德梧(2004)、张卫国(2006)等多数学者认为在对外汉语教学中汉字的量定为 3000 字左右比较合适。

从韩国的情况来看,韩国文教部早在 1972 年制定了《汉文教育用基础汉字》,共收录 1800 字。近年来汉文教育学会和文化观光部,先后向文教部提出修正案,所收汉字各为 2000 字,但未被文教部采纳,最后文教部根据自己的标准重新修订,并调整 44 个汉字,2000 年 12 月颁布新的《汉文教学用基础汉字》,其字数仍保持 1800 字。

据首尔大学许成道教授的 950 万字汉字语料的统计数据来看,1800 字左右的汉字在韩国各种书面材料中的覆盖率为 95%,而且 3000 字左右的汉字量,对韩国语料的理解度为 98%以上。(이준석, 2011) 这跟中国的情况基本相似。因此,我们认为《对韩字表》的字数总量定为 3000 字左右比较合理。这个汉字数在韩国留学生汉字使用频度统计中也得到了佐证。我们通过计算机软件对“多层偏误标注的国别化汉语中介语动态语料库”<sup>③</sup>中的韩国学生中介语语料进行检索,从 150 万字语料中统计出 3533 个汉字。在这 3533 个汉字中,超纲字有 787 字,而且超纲字中 80%以上是频次在 10 次以下的低频字,除去这些低频字,韩国留学生汉字使用字数也在 3000 字左右。

## 2 字表的选字原则

《对韩字表》是针对韩国学生的字表,因此我们不仅要考虑目标语,还要考虑学习者的母语情况,还要考虑最为第二语言的汉语汉字教学,因此字表必须在多维度下制定出来。据此,我们初步拟定以下四条收字原则:

1. 以中国的汉字字表为本,参照《汉语水平汉字等级大纲》,借鉴《基础汉字表》。
2. 以韩国汉字字表为辅,参照《汉文教育用基础汉字》,参照韩国《全国汉字能力检定试验》考级汉字。
3. 以“多层偏误标注的国别化汉语中介语动态语料库”研究成果作为维度,对汉字的筛选及等级划分进行检验和敲定。
4. 对中韩两国典型地名、姓名用字以及生活中个别带有文化因素的字进行专家干预。

《汉语水平汉字等级大纲》的选字是从频度统计角度、语言学角度、对外汉语教学角度、学生语言习得角度等四个角度出发,在常用性、均匀性、科学性、实用性、联想性、包容性、序列性八条原则下进行的。《汉字大纲》1992 年正式出版发行,历经 20 年为对外

<sup>②</sup>《汉语水平词汇与汉字等级大纲》,北京语言学院出版社,1992: 3。

<sup>③</sup>同①。

汉语教学做出了很大的贡献，但也存在一些问题，有待修订和补充。这方面周健先生已经做了很多工作和研究，他以《中国语言生活状况报告（2005）》中的数据为参照和依据，结合对外汉语教学实践，提出修订《汉字大纲》的意见并制定出《基础汉字表》。他对汉字大纲的选字原则归纳为以下四点：（1）《汉字大纲》的选字应跟常用词汇具有高度相关性。（2）《汉字大纲》的选字必须考虑汉字的构词能力。（3）《汉字大纲》的选字还要考虑到汉字自身系统建构的需要。（4）满足对外汉语需要的用字，如姓名地名等用字。此外，还有常用数量词、虚词等用字都应考虑收入。<sup>④</sup>我们非常认同周先生的观点，《对韩字表》的制定将充分借鉴《基础汉字表》。

韩国的《汉文教育用基础汉字》的初衷是为了进行汉文教育，所谓汉文指的是中国的古文献以及经典。虽然其所定目标是为了汉文教育，但根据国立国语研究院的汉字频度调查可以发现，这个汉字表不仅仅是为了汉文教育，不少国语教育用汉字也包括在里面。也就是说，虽然其名称是“汉文用”，但实际上带有双重性质，既包括汉文用汉字，也包括国语用汉字。但此字表所收录的汉字只有 1800 字，离 3000 字的字表汉字量有一定的差距。于是，我们又注意到韩国汉字等级考试所规定的汉字。韩国目前汉字等级考试主持机构繁多，其中唯一被政府认定的资格考试是“全国汉字能力检定试验”，该等级考试分 8 个等级加特级，每个等级所定汉字是根据韩国大规模语料的频度挑选出来的，其中第一级所要认读的汉字数为 3500 字，对我们有很大的参照价值。

“多层偏误标注的国别化汉语中介语动态语料库”是胡晓清教授主持的国家社科规划项目——“多层偏误标注的国别化汉语中介语动态语料库建设与研究”（11BY050）的研究成果。该语料库是胡晓清教授的山东省社会科学规划项目“韩国留学生汉语中介语语料库建设”（08CWJ09）为依托，开发建设的针对单一国别留学生的汉语中介语语料库，现已达 300 万字语料。该语料库提取库中的高频汉字建成《韩国留学生常用字频表》，同时，提取库中的偏误字，还建成《韩国学生偏误字表》。这些研究成果对《对韩字表》的建设提供有力的支撑和依据。

### 3 字表的等级划分

国内《汉字等级大纲》主要分为甲、乙、丙、丁四个等级，韩国基础汉字表分为初中用、高中用两个等级。周健先生提出“基础汉字、中级汉字、高级汉字”三个等级的建议，认为这与约定俗成的“基础汉语、中级汉语、高级汉语”提法相关联。我们比较认同此观点，至于哪些汉字收在哪个等级，具体收多少汉字还需要对中韩汉字进行多方面的比较和统计。

#### 3.1 韩国基础汉字在《汉字等级大纲》中的分布考察

下面我们先看看韩国基础汉字（1800 字）与中国大纲汉字（2905 字）之间的比较数据。

<sup>④</sup>周健（2007：212）。

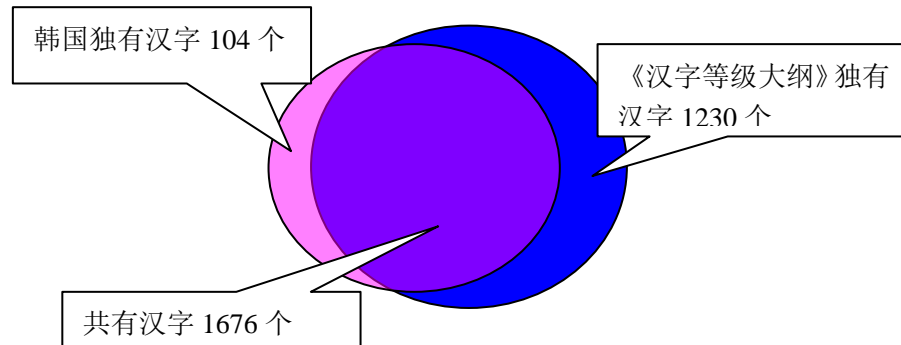


图 1

通过上图我们可以看出，韩国基础汉字表中韩国独有的汉字共 104 个，包括在《汉字等级大纲》中的有 1676 个<sup>⑤</sup>，占总数的 93.11%，占《汉字等级大纲》汉字的 57.69%。其中的中韩共有汉字显然是对韩字表首先要收进来的部分，等级也应该靠前的汉字。

下面我们进一步考察这些共有汉字在《汉字等级大纲》甲、乙、丙、丁各级的分布情况：

表 1 韩国常用汉字与《汉字等级大纲》共有字在《汉字等级大纲》中的分布情况表

等级	初中	高中	合计	比例 1 (占韩国常用字)	比例 2 (占《汉字等级大纲》)
甲	522	171	693	38.5%	23.85%
乙	232	312	544	30.22%	18.73%
丙	76	195	271	15.06%	9.33%
丁	44	139	183	10.17%	6.30%

如上表所示，两者的共有汉字中，甲级字最多，依次是乙、丙、丁级字。如此看来，韩国的常用汉字与《汉字等级大纲》有很强的共性。这些数据是我们分等级时要参考的主要依据。

### 3.2 学生汉字使用情况

除了字表之间的比较数据，韩国留学生汉语中介语语料库的字频统计数据也是我们要考察和依据的对象。我们利用计算机软件 MATLAB 提取了韩国留学生 150 万语料中所出现的所有汉字 3533 字以及它们的使用频度并与《汉字等级大纲》进行了比较，情况如下图所示：

<sup>⑤</sup> 《中日韩常用汉字对比分析》提到“经核对，900 初中汉字和 900 高中汉字中一字重叠（球），故本资料 K 按 1799 字计算”，因此它统计出的共有汉字是 1675 个。而经作者核查，这重叠的“球”字，应该是打印错误，实际上一是“球”字，另一个是“求”字。因此我们统计的共有汉字为 1676 个，总数是 1800 字。（4 页）

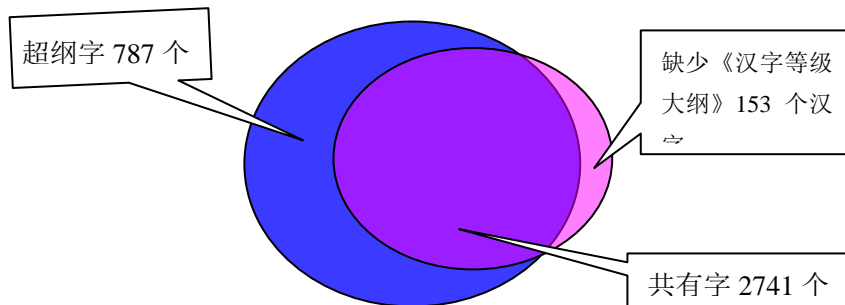


图2 韩国留学生汉字使用情况与《汉字等级大纲》汉字的比较

韩国留学生使用的汉字包括在《汉字等级大纲》中的有 2741 个，占《汉字等级大纲》的 94.73%，其中涵盖了甲乙两级汉字 1604 个，丙级字 566 个，丁级字 582 个。而韩国留学生实际使用的汉字中则缺少丙级字 35 个，丁级字 118 个。

另外，通过检索我们发现韩国留学生使用的汉字中超纲字有 787 个，其中有 69 个字是韩国 1800 个常用汉字中的汉字。超纲字的使用频度一般较低，使用频次在 10 次及以下的字占到全部超纲字的 80.94%。我们考虑分级时，这些因素都要考虑进去。

### 3.3 韩国留学生汉字使用频次分级

利用 MATLAB 软件我们还提取了韩国留学生汉字使用频次，并且按照使用频次的多少进行降序排列 A、B、C、D、E 五个等级，然后依据《汉字等级大纲》甲（800 字）、乙（804 字）、丙（601 字）、丁（700 字）各个等级的汉字数量将 3533 个汉字进行了等量的分级，依次记作 A（800 字）、B（804 字）、C（601 字）、D（700 字），超过 2905 个汉字以外的部分记作 E。3533 个汉字在《汉字等级大纲》甲、乙、丙、丁四级汉字中分布情况如下：

表2 韩国留学生所用汉字在《汉字等级大纲》中的分布情况

等级分段	甲	乙	丙	丁	超纲字
A (800)	603	170	17	9	1
B (804)	166	380	174	65	19
C (601)	24	175	170	143	89
D (700)	7	67	154	235	237
E (628)	0	12	51	122	443

通过对上表的分析，我们发现韩国留学生使用的汉字中 A 段与《汉字等级大纲》甲级字的共有汉字有 603 个，重合率为 75.38%；B 段汉字与乙级字的共有汉字有 380 个，重合率 47.26%；C 段汉字与丙级字相同汉字有 170 个，重合率为 28.29%；D 段汉字和丁级字相同的汉字有 235 个，重合率为 33.57%；另外 E 段汉字还包含了乙、丙、丁三级汉字中的 185 个汉字。从以上的分析数据可以看出，一方面，韩国留学生实际使用的 A、B 段汉字中，分别是甲级字与乙级字所占的比重最高，但是，C、D 段的汉字中，却分别是乙级字与超纲

字所占的比重最高；另一方面，韩国留学生使用的汉字按照使用频次高低依据《汉字等级大纲》各等级的字数进行分段后与甲、乙、丙、丁四个等级汉字的重合率并不高，除了和甲级字的重合率为 75.38% 以外，其他各级都在 50% 以下。

以上分析表明《汉字等级大纲》的制定只根据目的语（汉语）的字词频度来分级的情况与韩国留学生汉字实际使用频度并不相符，因此制定对韩汉字教学用字表时我们要根据韩国学习者的实际情况进行调整。

## 4 字表的标注内容

汉字的形、音、义是汉字的主要属性，但国内《汉字等级大纲》只列出汉字，也就是字形，对字的音和义都未提及到。对现有字表的简洁已有不少学者提出建议，下面我们结合日本和韩国的字表具体比较分析一下。

### 4.1 字形标注

中国《汉字等级大纲》的汉字字体是“宋体”，韩国汉字字表的字体是“明朝体”，日本的《常用汉字表》也用“明朝体”。对字体的选择，只有日本的字表作了解释，它介绍到“为了便于使用只采用现代通用印刷体——明朝体这一种字体用例”，而且对字体做了进一步的解释。

首先介绍各种版本明朝体的细微差异，比如，部首偏旁组合上的差异、笔画组合上的差异、笔画性质上的差异、特定字种之间的差异等，并举具体例字进行详细介绍。

其次，介绍明朝体与作为书写体（即笔记体）楷体之间的关系。这一部分也举具体例字进行了分析。

最后，介绍书写体楷体与印刷体不同带来的字体的不同。有些人书写时模仿印刷体，因此有些汉字跟楷体在书写上表现出不同的字形来。但字表认为，书写时采用哪一种问题不大，都可以用，但它指出哪种书写形式是更常用的书写形式。

整体来说，日本的常用字表做得非常细致，给使用者提供比较详细的参考和标准，相比之下，中国和韩国对此没有任何解释和说明，说明这方面重视不够。

下面列出在汉语里同一个汉字在宋体和楷体不同字体中显示的字形差异：

表 3

宋体	月	田	雨	质	垂	茶	进
楷体	月	田	雨	质	垂	茶	进
	两个横到不到头	中间的十到不到边	上面横的长短	十字的形状	中间左右两个竖的形状	木字撇点的形状	偏旁之的形状

对外国学生来说，类似于上面一些字的差异是很大的，中国人已经司空见惯，觉察不出来，但对于开始学习汉字的初学者来说，这些细节是需要讲解的。从学生的汉字偏误语料中，我们可以发现不少偏误是跟字体有关，学生没有掌握好不同字体中笔画形状的不同，

而且我们的教学也没有提及这部分，再加上高年级教师有时用行楷。学生遇到一个汉字的形状时无所适从，往往胡乱画过去，久而久之，学生慢慢不讲究细节，导致书写上的偏误。

总之，字形标注应该考虑字体，用宋体还是用楷体，也就是要印刷体还是书写体需要考虑进去。另外，不管采用哪一种字体，必须解释和说明因字体不同而造成的字形之间的细微差异，给学生提供借以参考的依据和标准，这方面可以借鉴日本的常用字表。

#### 4.2 字音标注

国内《汉字等级大纲》没有标注字音，而且没做任何解释，字音主要体现在词表上。韩国的字表标注了字音，而且只标注代表音，比如“乐”有两个音，【ak】(yuè)、【rak】(lè)，字表把【rak】音作为其代表音，并在前言中提到，“关于一字多音的汉字，字表只标一个字音，但实际教学中其他音也要教”。日本汉字有音读和训读，所以基本上都是一字多音，因此日本的《常用汉字表》列出了汉字的音，有些不规则音列在备注栏上。

一字多音是汉字当中比较常见现象，《对韩字表》要不要标注一字多音，如果标注，要标注几个音，都标注哪些音等等都是需要推敲的问题。

#### 4.3 字义标注

《汉字等级大纲》没有字义标注，韩国汉字字表也没有标注字义，但教学中教授代表训，也就是代表义。而日本的字表虽然没有对字义进行标注，但列出了例词。比如，

表 4

汉字	音训	例	备注
亜	ア	亜流，亜麻，亜热带	
愛	アイ	愛情，愛讀，恋愛	愛媛（えひめ）県
悪	アク オ わるい	悪事，恶意，醜恶 恶寒，好恶，憎恶 悪い，悪さ，恶者	

汉字的字义标注是比较复杂的问题，越是常用汉字其字义越丰富义项越多，如何标注字义确实是比较棘手的问题。可是日本常用字表列出例词的方法比较可取，这种方法既解决多音标注问题，也可以解决一字多义的问题，值得我们借鉴和应用。

## 5 字表在教学中的应用

关于汉字，大家的关注点在于汉字习得与汉字教学，对字表本身没有给予足够的关注。我们认为，汉字教学也好，汉字习得也好，必须有一个学习者和教学者所依据的基准。字表不能只列出汉字字种了事，它不仅要规定学习者在不同学习阶段所要学习的汉字的量，还要列出汉字的三大属性——即字形、字音、字义。字表在教学中的应用，我们可以从以下四个方面进行探讨。

第一, 字表必须分出汉字的级别来, 它是不同程度的汉语学习者在不同学习阶段所要掌握的汉字。我们可以根据每个等级所收录的规定汉字编写教材, 严格控制字量, 先教频度高、构词能力强的汉字, 循序渐进。

第二, 字形是文字最重要的属性, 从中介语语料库中的汉字偏误来看, 字形偏误占很大比重, 是不可忽视的问题。字表要对汉字字形的书写体和印刷体进行说明和分析, 并且让学生了解汉字的不同变体, 对汉字有全面的了解和认识, 避免不必要的书写错误。学习者和教师都要以字表所列的汉字字形为参考和依据, 尤其是教师应严格按照字表的标准汉字字形来规范自己的板书, 避免教师的随意板书给学生带来的困惑。即使学生遇到困难, 只要字表能提供一个规范而标准的字形, 学习者可以自行解决, 而且因不同字体带来的困惑会减少许多。

第三, 一字多音是汉字不可忽略的问题, 字表要显示汉字的不同音, 到底标注哪个音, 标注几个音, 都是字表要斟酌的问题。一旦字表标注了汉字音, 学习者可以根据字表, 了解每个汉字的不同语音, 教师也根据学生不同阶段有选择性地教学一个汉字的不同音。

第四, 绝大部分汉字是一字多义, 字表要挑选汉字最基本、最常用的字义。这不仅对学习理解词语的意思、了解汉语的词语结构带来方便, 对提高造词能力有很大的帮助。

总之, 字表是一个标准, 是一个依据, 它对汉语教学和习得都具有深远的影响。

## 参考文献

- [1]한문 교육용 기초한자 1,800 자 조정 백서. 교육인적자원부, 2000.
- [2]김광해. 등급별 국어교육용 어휘. 서울: 도서출판박이정, 2003.
- [3]김영옥. ‘한문교육용 기초 한자’ 제정에 대한 사적 고찰. 2012년도 한자한문연구소 동계학술대회, 2012.
- [4]이준석. 한문교육용 기초한자 문화관광부 재조정안에 대해. 2011.
- [5]진재교. 현행 한자한문 급수제도의 문제와 개선방안. <한문교육연구>, 한국한문교육학회. 2002.
- [6]白川静. 常用字解, [第二版]. 日本: 平凡社, 2012.
- [7]財前謙. 新常用汉字 196 ホントの書きかた. 日本: 芸術新聞社, 2013.
- [8]久米公. 漢字指導の手引き. 日本: 教育出版株式会社, 2013.
- [9]常用漢字表(平成22年11月30日内閣告示). 日本内閣, 2010.
- [10]新常用漢字表の文字論. 文字研究会. 日本: 勉誠出版株式会社, 2009.
- [11]野崎邦臣. 漢字字形問題点. 日本: 天来書院, 2013.
- [12]中日韩常用汉字对比分析. 北京书同文数字化技术有限公司. 商务印书馆, 2009.
- [13]“外国学生错字别字数据库”的建立与基于数据库的汉字教学研究. 北京语言大学“外国学生错字别字数据库”课题组. 语言教学与研究, 2006(4).
- [14]陈 纭. 日本学生书写汉语汉字的讹误及其产生原因. 世界汉语教学, 2001(4).
- [15]甘瑞瑗. “国别化”对外汉语教学用词表制定的研究. 北京: 北京大学出版社, 2006.
- [16]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. 北京: 北京语言学院出版社, 1992.
- [17]黄溢芝. 对外汉语教学汉字字表统计研究. 语言教学研究, 2010(4).



- [18]刘金珠. 论汉语水平汉字等级大纲的研制. 暨南大学硕士论文, 2010.
- [19]刘丽媛. 基于韩国留学生汉语中介语语料库的汉字偏误研究及应用. 鲁东大学硕士论文, 2013.
- [20]刘英林、宋绍周. 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: 北京语言学院出版社, 1992.
- [21]张 凯. 汉语构词基本字的统计分析. 语言教学与研究, 1997 (1) .
- [22]李清华. 《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, 1999 (1) .
- [23]姜德梧. 关于《汉语水平词汇与汉字等级大纲》的思考. 世界汉语教学, 2004 (1) .
- [24]张卫国. 阅读: 覆盖率、识读率和字词比. 语言文字应用, 2006 (3) .
- [25]王 艳. 基于韩国留学生汉语中介语语料库的汉字偏误研究. 鲁东大学硕士论文, 2011.
- [26]杨 林. 对外汉语教学《初级汉字表》研制. 北京语言大学硕士论文, 2011.
- [27]周 健. 汉字教学理论与方法. 北京: 北京大学出版社, 2011.