

文章编号: 1003-0077 (2011) 00-0000-00

对外汉语教学领域话题语料库的研究与构建*

胡韧奋¹, 朱琦¹, 杨丽姣¹

(1.北京师范大学, 北京市 100875)

摘要: 对外汉语教学领域, 教材上的课文通常围绕一个话题展开, 话题是教学内容的集中的体现, 也与词汇、语法等不同层面的语言知识间有着密切关联。该文基于大规模教材语料库研究教学话题分类体系, 设计了一个包含 4 个一级话题、23 个二级话题和 246 个三级话题的三层话题框架, 并据此对 197 册汉语经典教材中的 5457 个文段进行了人工标注及校对, 构建了一个规模约 12 万句的面向对外汉语教学的话题语料库。为了更好地服务于汉语教学及相关研究工作, 还抽取、计算了文段的语法点和新 HSK 词语等级信息, 作为话题标注的补充维度加入资源库, 以为汉语教学领域的教师、研究者及教材编写者提供较为全面的话题信息参考。

关键词: 对外汉语; 话题; 语料库

中图分类号: H087/ H195

文献标识码: A

The Research and Construction of a Topic Corpus for Teaching Chinese as a Second Language

Hu Renfen¹, Zhu Qi¹, Yang Lijiao¹

(1.Beijing Normal University, Beijing, 100875, China)

Abstract: In the area of teaching Chinese as a second language, each text in the textbook has a specific topic. Topic represents the core contents of each lesson, and has close relationship with other linguistic knowledge such as vocabulary and syntax. This paper introduces a hierarchical topic bank with 4 level-1 topics, 23 level-2 topics and 246 level-3 topics. The authors manually labeled 5457 texts from 197 classical Chinese textbooks based on the topic bank and built a topic corpus including over 120 million sentences. In order to offer comprehensive reference on topic information, syntactic constructions and HSK word level information are also extracted as supplement knowledge for topic labeling.

Key words: Chinese as a second language; topic; corpus

1 引言

话题是文本内容的体现, 也是衔接语言和文化的重要桥梁。在对外汉语教材中, 课文通常围绕一个话题展开, 语言技能和文化知识以该话题为语言环境进行组织和呈现, 从而达到教学目的。关于话题在汉语教学中的重要性, 朱志平等(2008)在对 548 种对外汉语教材进行调研后提出, 话题是教学内容设计思想的集中体现。李亚丽(2012)指出, 话题的选择是课文编排的重要内容之一, 直接关系到学生综合语言运用能力的培养。课文话题的选择不但决定着常用词语的选择取向, 而且影响着语法结构、功能项目及文化项目的融入等。

国家汉办出版的《国际汉语教学通用课程大纲》给出了《汉语教学话题及内容建议表》, 该表列有以下 22 类话题, 是教材编写及课堂教学中话题选择与编排的重要参考:

个人信息、情感与态度、社会交往、日常生活、学校生活、家庭生活、文化娱乐、节日活动、身心健康、习俗与禁忌、旅游与交通、语言与文化、价值观念、文学与艺术、政治、历史与地理、科学与技术、全球与环境、计划与未来、热门话题、教育、植物与动物、自然景观

* 收稿日期:

定稿日期:

基金项目: 国家语委“十二五”科研规划项目“语言资源建设规划研究”(编号: YB125-124)

为了帮助教材编写者和汉语教师更好地把握话题设计,苏新春(2011)等对部分教材中的话题分布情况进行了调研和统计,余千华(2012)、罗洋天(2013)、李欣蓓(2014)等则从学习者出发,通过问卷和访谈调查学习者的话题兴趣及其与教材课文的匹配情况,进一步丰富了话题选择与编排的参考依据。然而,现有研究也存在以下局限因素:(1) 研究范围限于特定级别(如汉语中级水平)或特定类型(如口语)教材,未能提供全面的参考和参照;(2) 研究中,话题的选取和分类体系缺乏大规模语料的验证,具有一定的主观性。

为了避免主观因素影响,向不同等级、不同类型的教材编写及汉语教学提供话题参考依据,本文基于大规模教材语料库开展话题类型研究,对197册经典对外汉语教材中的课文进行调研和标注,以此为依据制订较为综合和全面的话题标注框架,该框架设置三级节点,包含4个一级话题、23个二级话题和246个三级话题。标注工作面向教材语料库中的5457个文段展开,每个文段依次被标注一级话题、二级话题和三级话题,例如《跟我学汉语》第二册第7课《给你一张电影票》第一个文段(如下)在语料库中被标注为“一级话题:日常生活;二级话题:娱乐休闲;三级话题:看电影”。

安妮:小雨,你想不想看电影?

小雨:什么电影?

安妮:《花木兰》。

小雨:太好了!今天的还是明天的?

安妮:今天下午4:30的。刘老师给我两张票,我给你一张。

小雨:今天不行,我没空儿。

安妮:你要干什么?

小雨:我要做海报。你给马明吧!他是个影迷。

安妮:好吧。再见!

经过人工标注及校对,我们构建了一个面向汉语教学的综合话题信息语料库,以期为不同水平、不同类型的汉语教学及教材编写提供参考。语料库中的文段除了标注详细的话题信息外,还标注了该文段包含的语法点及HSK词语等级信息,以帮助汉语教师及教材编写者更好地选择和使用库中资源。

2 语料来源

语料的选择对于面向汉语教学的话题语料库建设而言十分重要,因为语料既是标注的对象,又是标注体系(即话题类别体系)制订的参考和依据。从汉语教学的实际需求出发,我们将经典对外汉语教材文本作为主要研究对象。首先从汉语国际教育动态语料库¹中提取了122458句对外汉语教材语料,共计5457个文段及其来源教材、单元、课文名称等信息。

所选取语料来源于197册经典对外汉语教材。动态语料库在采集教材时充分考虑了教材类型、适用水平、出版年代、影响因子等属性特征(杨丽姣,2015)。话题信息的标注结果将融入汉语国际教育动态语料库,作为其中一项功能供用户检索和使用。

3 话题分类研究及标注框架设计

为了制订科学合理的话题标注框架,除了《国际汉语教学通用课程大纲》中给出的《汉语教学话题及内容建议表》外,我们也对汲传波(2005)、苏新春(2011)、李亚丽(2012)、刘华(2014)等人关于话题分类的研究进行了细致考察,认为框架设计应充分考虑话题的层次性、内容的涵盖面、类与类之间的区别度、每个节点的重要程度与颗粒度等因素。基于以

¹ 汉语国际教育动态语料库由北京师范大学中文信息处理研究所和汉语文化学院共同建设,主要收录对外汉语教学领域经典教材和新HSK样卷文本语料,并提供多层次的语言信息标注,目前规模约14万句,240余万词次。

上考量，我们设计了一个包含三个层次的话题节点表作为标注框架，参见表 1。

表 1. 面向对外汉语教学的话题节点表

一级节点	二级节点	三级节点
基本信息	人物	人物信息（年龄、国籍、职业……）；家人与朋友；外貌（相貌、衣着）；身体部位；爱好与特长；性格；其他
	事物	天气与气候；日期；时间；数字；方位；物品属性（数量、颜色、大小、归属、品牌…）；其他
日常生活	餐饮	食物；饮料（茶、咖啡…）；就餐（约餐、订餐、用餐…）；点餐；做饭（买菜、洗菜、做菜…）；餐具；菜谱；其他
	出行	问路；交通方式；票务；住宿事务；出入境；旅游；其他
	购物	买东西（描述/比较商品、询价与砍价）；退货与换货；商品质量；其他
	娱乐休闲	读书；玩游戏；听音乐；看电影；看电视；看演出；看比赛；养宠物；养植物； 摄影；下棋打牌；唱歌跳舞；逛街出游；参观(博物馆…)；其他
	时尚	追星；手机通讯；美容与整容；网聊；网购；拼客团购；名牌；流行音乐；其他
	运动与健康	运动；健身；运动会与比赛；生理健康：受伤；生病；看病；吃药；探病；生活习惯；养生；减肥；怀孕生子；季节变化与生活关系；心理健康：心灵鸡汤；情感/情绪；其他
	家庭与社区	亲情；亲子关系；家庭教育；婚姻生活；社区生活；其他
	日常事务	作息；穿着打扮；打扫卫生；洗衣；洗照片；环境布置；维修；搬家；理发；收/发快递（信件）；借/还物品；办证/卡；计划；意外（迟到、车祸…）；报警；接机；送机；生活经验及常识；丢/还物品；留言；其他
社会生活	学习与成长	学汉语；课程与选课；作业与研究（请教/讨论…）；请假；社团及其他校园活动；周末假期；考试；游学/留学；选择学校/专业；奖学金；毕业；校园食宿；学习压力与困难；理想；人物经历；成长烦恼；其他
	人际交往	拉家常；友谊；恋爱择偶（追求、相亲…）；师生关系；邻里关系；同事关系； 约会（同学、朋友、恋人等相约）；聚会（生日晚会…）；请客；做客；送礼；家长里短；其他
	职场	职业介绍；个人层面：求职；职业选择；福利待遇；职场生存；加班；工作日程；开会；兼职；出差；休假；创业；晋升；辞职；企业层面：介绍企业/产品；招聘；培训；裁员；其他
	经济生活	个人层面：银行业务（如换钱…）；租房；买房；买车；金融生活（信用卡、投资理财…）；投诉索赔；公司层面：公司注册；法律事务；产品创新与知识产权；招商加盟；广告；营销；谈判；进出口贸易；物流；投资；售后；其他
	社会现象	交通问题（拥堵、事故…）；犯罪；法律纠纷（打官司、判刑…）；社会公德；就业；失业；环保/污染；经济发展与社会变化（春运、春晚、留守儿童、养老、人口问题、贫富差距、消费者权益…）；婚姻家庭问题；教育（高考、育儿…）；网络问题；医疗健康；食品安全；男女平等；名人与偶像；公益；新能源；城市建设；新闻与传媒；时事政治；公共外交；其他
文化生活	语言文化	名字与称谓（人名、地名、称呼…）；汉语汉字（谐音、对联、语言避讳、汉字文化…）；成语/俗语故事；普通话与方言；笑话；学汉语的经历及故事；

		其他
	文艺作品	文学作品（寓言、散文、童话…）；艺术作品及赏析（音乐、戏剧、摄影、影视…）；创作背景/缘起；其他
	文化习俗	节日（春节、圣诞…）；传统习俗与礼俗（如婚礼…）；礼仪（课堂、餐桌、待客…）； 生活习俗（衣食住行的习惯和风俗）；其他
	文化产物	中国茶；中国功夫；生肖与星座；中华饮食；建筑艺术；中医；书法国画；戏曲； 四大发明；手工艺（如剪纸、风筝…）；相声曲艺；服饰文化；棋牌；其他
	文化观念	人生哲理；宗教文化；儒家道家思想；民间百科；其他
	多元文化	异文化印象；异文化适应；文化比较；文化禁忌；移民（如海外华人故事…）；其他
	地理与历史	地图；景观/景点；城市地区（城市比较、介绍故乡…）；地理与文化；五十 六个民族；历史广角；历史人文故事；其他
	自然与科学	动物；植物；人与自然；小百科；科技发展；探索与发现；科学幻想；其他

除了语言技能（如词汇、语法）的传授外，语言在生活中的运用也是教学中的重要关注点，故而，话题的选择应尽量贴近生活。一级话题的设计从物质生活和精神生活两个维度出发，物质生活可落实到“日常生活”和“社会生活”这两个节点，精神生活在教学中主要表现为文化元素的融入，故而我们设置了“文化生活”节点。此外，在汉语学习的初级阶段，大部分内容是关于人物、事物基本信息的讨论，故设置“基本信息”这一节点。四个一级话题按顺序排列为：基本信息、日常生活、社会生活、文化生活，这体现了与学习者个人关联的远近和内容层次的提升。

二级话题的设计上，“基本信息”所包含内容较为明确，可分为“人物”和“事物”两个二级节点。“日常生活”和“社会生活”下设二级节点的设计则主要基于教材中常出现的单元主题，并综合考虑学习者兴趣和话题的频度信息。“文化生活”的展开方式较为特殊，我们参考杨丽姣（2009）提出的汉语教学文化项目整理思路，借鉴美国《21 世纪外语学习标准》在文化教学上提出的“文化三角观”来对文化生活进行归类。所谓“文化三角观”，即文化包括一个社会所有的哲学观念、行为习俗和社会产物（有形的和无形的）。故而，我们在“文化生活”话题下设“文化习俗”、“文化产物”和“文化观念”节点，并结合教材中的课文选题增加了“语言文化”、“文艺作品”、“地理与历史”、“自然与科学”节点。此外，我们注意到跨文化交际是汉语教学中的一项重要内容，也是每个二语学习者在学习和使用汉语时都会面临的问题，因而增设“多元文化”节点作为一项二级话题。

三级话题在大纲及现有的话题分类研究成果的基础上，结合对不同类型教材课文的调研结果给出，并由专家作补充和完善。值得一提的是，话题节点表需要经过大规模实际语料的检验，因而我们通过语料标注来对已有节点表进行修订。首先，对语料库中的 2 万句教材语料进行试标注，并对节点表进行初步修订，一级话题和二级话题的框架在这一阶段基本得到确定。随后，在后期 10 万句人工标注中不断收集标注者的反馈建议，进一步修订节点表中的三级话题，从而得到表 1 中的分类结果。标注修订的方法综合考虑了节点的频次、颗粒度等信息。

4 话题信息的标注实践

为了保证话题信息的标注质量，我们组织了 9 名语言学及相关专业本科生、研究生参与人工标注，标注人员经培训及试标注检验合格后，进入正式标注。标注结果经人工校对，存疑处由专家讨论确认。如上文所述，话题节点表在标注中得到不断地检验和修订，具体工作

分为以下九个步骤：

- 步骤 1. 切分文段边界，确认话题标注单位²；
- 步骤 2. 为文段标注一级话题和二级话题；
- 步骤 3. 收集一级、二级话题的标注反馈并调整节点表；
- 步骤 4. 根据更新节点表修订标注结果；
- 步骤 5. 为文段标注三级话题；
- 步骤 6. 收集三级话题的标注反馈并调整节点表；
- 步骤 7. 根据更新节点表修订标注结果；
- 步骤 8. 人工校对所有标注信息；
- 步骤 9. 确认最终话题节点表和标注结果。

在步骤 2 和步骤 5 的话题信息标注中，标注者需注意对以下几种特殊情况的处理：

(1) 文段涉及多个话题：如果认为文段涉及两个话题（包括二级话题和三级话题）的情况，可为该段落标注两个话题，如果认为文段涉及三个及以上话题，则为该段落标注最为凸显的两个话题。

(2) 文段内部涉及话题跳转：如果文段前后围绕不同话题展开，比如《中文起步》第一册第 19 课《中国城》这篇课文为一段长对话，甲乙两人首先讨论了怎么去中国城（“日常生活-出行”），然后讨论了在中国城一家饭馆点餐吃饭的经历（“日常生活-餐饮”），在标注时需要给出段落原话题、跳转新话题及跳转句的语料编号。

(3) 文段内容无法按照已有的话题节点表进行标注：标注者在备注栏给出问题描述及对话题节点进行增加或修改的建议。对于其他标注者无法确定的话题标注信息，也可在备注栏给出问题描述及相关建议。

(4) 三种情况暂不做标注：①课文内容为简单句子，不构成文段（多见于入门阶段的汉语教材）；②课文为简单的 1-3 轮短对话，不围绕特定话题展开，而强调交际功能（比如问好、道歉等，同样多见于入门阶段的汉语教材），将在语料库的交际功能项目中得到标注；③课文内容时代性、政治性较强（例如以知青、承包责任制等为背景的短文或小说），不适用于作为当前汉语教学及教材编写参考。

在步骤 3 和步骤 6 中，我们收集并总结标注者在备注栏给出的问题及建议，并侧重对涉及以上四种特殊情况的文段进行考察，通过专家讨论确定标注方案，以对已有话题节点表进行调整。

标注结果存储为 XML 格式，以下为一个文段的话题标注示例：

```
<text id="24" name="新实用汉语课本_Volumn 1_Lesson 12_我全身都不舒服_Text1"
topic_1="日常生活" topic_2="运动与健康" topic_3="生病">
  <s id="283">丁立波：大为，你每天都六点起床去锻炼，现在九点一刻，你怎么还不起床？
</s>
  <s id="284">马大为：我头疼。</s>
  <s id="285">丁立波：你嗓子怎么样？</s>
  <s id="286">马大为：我嗓子也疼。</s>
  <s id="287">丁立波：我想，你应该去医院看病。</s>
  <s id="288">马大为：我身体没问题，不用去看病。</s>
  <s id="289">我要睡觉，不想去医院。</s>
```

² 语料库中的语料以句子为单位存储，并按照其在教材中出现的顺序排列，在为文段标注话题之前，首先要做段落边界切分，确认一个文段的起始句和结束句。

<s id="290">丁立波：你不去看病，明天你还不能上课。</s>
<s id="291">马大为：好吧。</s>
<s id="292">我去医院。</s>
<s id="293">现在去还是下午去？</s>
<s id="294">丁立波：当然现在去，我跟你一起去。</s>
<s id="295">今天天气很冷，你要多穿点儿衣服。</s></text>

5 标注数据分析及应用

按照上述步骤，我们对从语料库中采集的 5457 个文段进行了人工标注及校对，共有 5168 个文段被标注话题信息（涉及两个话题的文段在此计为两段）。话题信息的标注结果可以直接反映当前汉语教学中的内容选择和关注倾向，接下来，将对一级话题、二级话题和三级话题的标注数据分别进行统计分析，挖掘其中规律，并进一步探讨标注结果在汉语教学中的应用问题。

5.1 话题信息数据分析

一级话题共有“基本信息”、“日常生活”、“社会生活”和“文化生活”四个类别，其分布情况如图 1 所示。从标注结果来看，共有 402 个文段被标注为“基本信息”，数量较其他三个类别较少，主要是由于“基本信息”话题通常仅出现于初级阶段的汉语教材中。“日常生活”话题频度最高，达 1759，因为它与初级、中级汉语教学均息息相关，是这两个阶段的主体学习内容。“社会生活”话题涉及中级、高级汉语及商务、职场汉语等专门领域的汉语教学，属于该话题的文段数量仅次于“日常生活”，为 1671。“文化生活”多出现于中级、高级汉语教材，该话题下的文段数量为 1336。

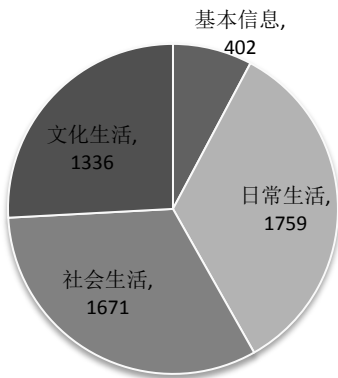


图 1. 一级话题数量分布情况

话题节点表中设有二级话题 23 个，其中，“日常生活”和“文化生活”各包含 8 个二级话题，“社会生活”下有 5 个二级话题，“基本信息”包含 2 个二级话题，这些二级话题在语料库中的数量分布情况如图 2 所示，其频度从高到低排列为：

社会现象、人际交往、出行、运动与健康、学习与成长、地理与历史、日常事务、语言文化、人物、工作职场、经济生活、家庭与社区、购物、娱乐休闲、餐饮、事物、多元文化、文艺作品、自然与科学、文化习俗、文化观念、文化产物、时尚

“社会现象”是最为高频的二级节点，这体现了教材对时事热点问题的关注，有助于学生在习得词汇、语法等语言知识的同时，对中国的社会现实背景有所了解。同时，可以注意到，除了“语言文化”外，“文化生活”下属的与中国文化直接相关的二级话题如“文化习俗”、“文化产物”、“文化观念”等在数量上并不占优势，说明当前的汉语教材在文化项目的整理和呈现方面尚有待进一步挖掘和提升。

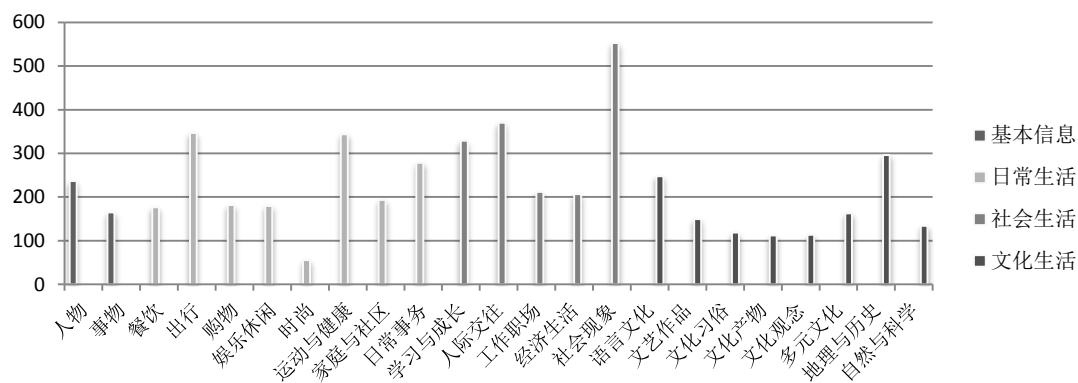


图 2. 二级话题数量分布情况

二级话题是对课文内容较为抽象的概括，近似单元主题，三级话题则较为具体，直接与课文（文段）内容对应。经过对教学内容的调研总结及语料标注实践，话题节点表在第三层给出了 246 个三级话题。我们对语料库中的三级话题标注结果进行统计，发现频度 ≥ 50 的节点有 25 个，从高到低排列如下：

买东西、经济发展与社会变化、人物信息、旅游、看病、历史人文故事、文学作品、交通方式、友谊、汉语汉字、恋爱择偶、文化比较、节日、方位、人生哲理、家人与朋友、问路、亲子关系、异文化印象、成语/俗语故事、点餐、景观/景点、运动、亲情、天气与气候

频度最高的节点是“买东西”，共有 172 个文段被标注该话题，这与学习者的实际使用需求密切相合。其他频度较高的三级节点分布在不同的二级话题及一级话题下，反映了教材中较为典型的课文内容。

5.2 话题语料库的实践应用

通过较为系统的调查研究及大规模的语料标注实践，我们构建了一个汉语教学话题节点表和标注了三级话题信息的语料库。资源的开发需要面向实际的需求，为了更好地服务于对外汉语教学及相关研究工作，我们对该领域的用户需求进行分析，并据此设计了基于话题信息的综合查询功能。

话题标注对象虽然是语段，但汉语教学却覆盖了语言的各个层次：词汇、短语、句子、段落等。因此，我们在设计话题信息的查询功能时，需要充分考虑不同层次的语言教学要点。杨寄洲（2003）在研究教材编写问题时指出：课文编写的最大难点除了生词量的控制和常用词出现的比例问题以外，还有个语法点与课文内容相协调相匹配的问题。由此可见，对于教材语料来说，语段话题与其包含的词汇、语法点之间的关联也是汉语教学的关注点之一。基于以上考虑，我们从汉语国际教育动态语料库语法点标注项目（谭晓平等，2015）中提取了句子包含的语法点信息，并自主开发 HSK 词语等级计算工具，以自动计算每个文段所包含的各等级 HSK 词语数量及比例。国家汉办发布的《新汉语水平考试大纲》中列出了 HSK 一级至六级词汇大纲，我们分别读取 HSK 一级至六级词汇表，并基于语料分词和词性标注结果进行词语匹配。计算各等级词语比例时不计标点、英文、数字和姓名。

通过语法点信息的抽取和词汇等级计算，我们将语段的词汇、语法点和话题信息进行了整合，实现了以话题为基础、辐射词汇和语法层面的资源建设。因而，用户在进入话题信息在线查询系统后，可以在话题列表中点击一个节点，以获取被标注该话题的语段，并得到语段的详细信息，包括语段内容、语段来源、语法点、新 HSK 词语等级信息等，基本体例如图 3 示。

话题：日常生活 - 餐饮 - 做饭			
段落来源：跟我学汉语 Volumn 3 Lesson 17 我把菜谱带来了 Text2			
<p>你会做中国菜吗？如果你想学习做中国菜，你应该先找到一个菜谱，然后把需要的原料买来，把它们洗干净。洗好以后，按照菜谱把原料切好，准备好作料再开始做。中国菜重视色、香、味。例如“鱼香肉丝”，有漂亮的红色和黄色，鱼一样的香味和又酸又辣的味道。你想不想尝尝？</p>			
语法点信息			
……如果…… 先……然后…… 把字句 按/按照 又……又……			
新 HSK 词语等级信息：			
等级	数量	比例	词语列表
HSK-1	28	39.44%	你会的有漂亮和学习想吗中国做好菜
HSK-2	6	8.45%	开始再鱼洗准备
HSK-3	13	18.31%	需要先吧然后以后干净一样又如果应该
HSK-4	7	9.86%	例如辣重视味道香酸按照
HSK-5	2	2.82%	原料

图 3. 话题查询界面示例

6 总结

话题是教材文本内容的体现，也与教学中的其他语言知识密切相关。本文基于大规模教材语料库研究教学话题分类体系，设计了一个包含 4 个一级话题、23 个二级话题和 246 个三级话题的三层话题框架，并据此对 197 册经典对外汉语教材中的 5457 个文段进行了人工标注及校对。为了更好地服务于汉语教学及相关研究工作，我们从用户实际需求出发，抽取并计算了文段的语法点信息和新 HSK 词语等级信息，以期为汉语教学及研究提供较为全面的话题信息参考。

基于上述研究工作，我们希望能从以下几个方面做出尝试，以改进和提升现有的资源，并探索新的应用空间：第一、将话题语料库及综合查询功能部署上线，广泛地收集用户使用意见，以修订和改善现有的资源及功能设计；第二、随着时代的发展和国力的增强，对外汉语教学的内容需求仍在不断地增长和变化，在现有的标注框架及资源的基础上，我们需要持续关注社会现象、流行时尚等开放话题，选择合适的新闻媒体报道作为补充语料，同时，需侧重对文化项目的整理和挖掘，除了中国传统文化外，中国的现当代文化、其他国家的文化也应作为节点设计和资源补充的重点关注对象；第三、开展话题关键词提取、话题自动识别等技术的研究，以丰富现有资源库建设维度，拓展资源的应用空间。

参考文献

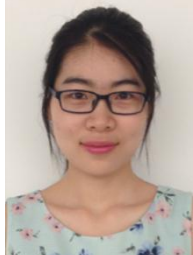
- [1] 朱志平, 江丽莉, 马思宇. 1998-2008 十年对外汉语教材述评[J]. 北京师范大学学报(社会科学版), 2008, 05:131-137.
- [2] 李亚丽. 对外汉语中级教材课文话题分析[J]. 大众文艺, 2012, 03:263-264.
- [3] 孔子学院总部, 国家汉办. 国际汉语教学通用课程大纲[S]. 北京: 北京语言大学出版社, 2014.
- [4] 苏新春, 唐师瑶, 周娟等. 话题分析模块及七套海外汉语教材的话题分析[J]. 江西科技师范学院学报, 2011, 06:58-65.
- [5] 余千华, 樊葳葳, 陈琴. 汉语学习者话题兴趣及其与对外汉语教材话题匹配情况调查研究[J]. 语言教学与研究, 2012, 01:23-29.
- [6] 罗洋天. 《匈牙利汉语课本》课文话题设计[J]. 云南师范大学学报(对外汉语教学与研究版), 2013, 06:58-68.
- [7] 李欣蓓. 从对外汉语教材话题的选择看编写者文化态度——基于三部对外汉语教材话题的分析[J]. 云南师范大学学报(对外汉语教学与研究版), 2014, 05:25-32.
- [8] 杨丽姣, 肖航. 面向语义搜索的语料库语境信息标注研究[J]. 语言文字应用, 2015, 01:107-116.
- [9] 汲传波. 对外汉语口语教材的话题选择[J]. 云南师范大学学报, 2005, 06:14-18.
- [10] 刘华, 方沁. 汉语教学用话题库及话题分类影视资源库构建[J]. 世界汉语教学, 2014, 03:378-392.
- [11] 杨丽姣. 文化项目的归类与国别化汉语教材的话题选点[A]. 厦门大学汉语国际推广南方基地. 2009年汉语国别化教材国际学术研讨会论文集[C]. 厦门: 厦门大学汉语国际推广南方基地. 2009:6.
- [12] American Council on the Teaching of Foreign Languages. Standards for Foreign Language Learning: Preparing for the 21st Century[S]. Kansas, USA: Allen Press Inc, 1999.
- [13] 杨寄洲. 编写初级汉语教材的几个问题[J]. 语言教学与研究, 2003, 04:52-57.
- [14] 谭晓平, 杨丽姣, 陈颖. 基于语料库的汉语语言点标注研究[J]. 曲靖师范学院学报, 2015, 02:103-108.
- [15] 孔子学院总部. 新汉语水平考试大纲(HSK 一级至六级)[S]. 北京: 商务印书馆, 2009.

作者简介:

1. 胡韧奋(1988—), 女, 博士研究生, 主要研究领域为中文信息处理, Email: irishere@mail.bnu.edu.cn;



2. 朱琦(1990—), 硕士研究生, 主要研究领域为中文信息处理, Email: azhuqi@126.com;



3. 【通讯作者】杨丽姣(1973—), 女, 副教授, 主要研究领域为汉语国际教育、词汇语义学、语料库语言学, Email: yanglijiao@bnu.edu.cn.

