

# 语言资源与智能

2018年第2期（总第7期）



语言资源高精尖创新中心 主编

2018年5月

## 目录

### 智库研究

- 语言规划研究的新形势（李宇明教授）

### 中心动态

- 语言资源高精尖创新中心召开学术委员会、咨询委员会会议
- 语言资源高精尖创新中心在日本成功举办首届“一带一路”语言资源与评测研讨会
- 法国著名汉学家白乐桑教授受聘为语言资源高精尖创新中心国际学术顾问
- 澳门理工学院代表团访问语言资源高精尖创新中心
- 针对成人学习者的汉字教学法研究——“语言资源与智能”大讲坛第十六讲

### 项目进展

- 《汉学研究》资源建设项目简介（阎纯德教授）
- 中华汉字文化体验园项目正式启动
- 中俄经贸合作数据库及俄汉语智能化综合服务平台项目顺利完成定期检查
- 中国周边（6国）语言资源库建设项目工作研讨会顺利举行
- 面向智能语音教学的汉语中介语语料库建设项目进展报告会暨工作研讨会顺利举行
- HSK 动态作文语料库 2.0 版重磅发布

## ► 智库研究

### 语言规划研究的新形势

李宇明教授

中国语言政策和语言规划研究面临着四个方面的新形势：

#### 第一，全球治理。

过去讨论语言规划，没有“全球治理”的概念，甚至连“治理”这个词也很少用。“全球治理”概念提出，是“一带一路”倡议的逻辑延伸。“一带一路”倡议是要建立“人类命运共同体”，将中国的利益和全人类的利益结合起来。要实现“一带一路”倡议，要建立“人类命运共同体”，就需要有全球治理的理念。我们国家的语言文字工作，我们的汉语国际传播事业，都需要上升到全球治理的层面来思考、来规划、来推进。

全球治理是全世界共同参与的治理。要实现全球治理，中国首先要把自己的事情办好。治理应该是“自上而下”和“自下而上”的结合，应该是社会利益团体之间的共商共享、共出智慧、共同发展。搞好我们自己国家的政治、经济、文化和意识形态，在国际社会中取得领先优势，才能谈到中国文化对人类共同体的贡献，才能贡献出充满中国智慧的全球治理理念。在这种大视野下，看中国的语言政策和语言规划，就会明确我们努力的方向，不是简单的查缺补漏，不是简单地增进对国际社会的了解，而是要明确中国在推进全球治理中的历史使命，明确中国的语言政策和语言规划如何服务于这一历史使命。

#### 第二，文化问题。

文化问题有两个方面。一方面是全球化带来的文化积聚、文化矛盾和文化冲击。文化多样性使得每个人的文化负担、每个社会群体的文化负担都在加重，这就要求我们，能够在不同的文化中穿行，能够应对可能的文化冲突，能够促进不同文化间的沟通交流。另一方面是传统文化的复兴。通过复兴传统文化来防御外来文化的负面影响，同时也要警惕文化复兴演变成民粹主义，目前世界上不少国家都出现了类似民粹的危险倾向。

文化冲突、文化多样性和传统文化复兴对教育领域的语言政策有什么影响？对汉语国际传播有什么影响？给语言政策和语言规划研究带来什么样的挑战？这是必须思考的。

### **第三，信息化问题。**

信息化是非常重大的问题。人工智能，特别是涉及语言的人工智能快速进入到我们的生活领域。这促使我们去思考两个核心问题：一是因果关系是不是还是最重要的关系？自从亚里士多德以后，人类科学信仰都是基于因果律的，我们相信一切东西都是有因果规律或因果关系的。但随着信息化的发展，这一切似乎正在改变。比如藉由人工智能算法推送给你的头条新闻或信息，就不是根据因果关系，而是相关性。所以有很多人在写文章讨论，因果关系和相关性哪个最重要。二是规则和知识还重要吗？比如阿尔法狗战胜了人类优秀棋手，但它不是基于对人类棋谱的学习，而是自己去重新发现围棋的规律。语言信息处理也存在类似情况，语料库语言学诞生之后，好多人已经不相信语言学了，产业界甚至有“赶走一个语音学家，计算机语音处理的水平就提高一个层次”的说法。深

度学习理论出来之后，所谓语法规则也基本上不是必要的了，因为机器自己会发现，而且机器的翻译水平的确越来越高。

更重要的是随着人工智能的发展，机器人将以更快的速度、更广泛地进入人类生活。此前的机器人，大都工作在比较困难的地方，比如自然条件异常恶劣之地、火场救灾、地下管道等，以后机器人会和人类并肩工作，特别像导医、导购、解说员等工作角色。

作为语言学工作者，我个人非常支持人工智能的发展，但我必须要说，面对人工智能的发展，语言学界有点失语，特别是在语言信息化方面失语，在促进语言信息发展的语言政策和语言对策上失语。比如，机器翻译已经可以满足人们对英语甚至是“小语种”日常交流的需求，那么，我们“全民学外语”是强调素质教育还是培养外语能力？外语专业是培养外语专家还是行业翻译？这都是语言政策和语言规划要研究的问题。

信息化为我们构造了一个虚拟空间，但这个虚拟空间的语言政策，一直是套用现实空间的语言政策。虚拟语言生活远比现实语言生活有创造力，用现实空间的语言政策不一定能管得住，管得好。此外还要注意，很多人正在被信息边缘化。现代语言技术，很多老年人不会用，很多学生不让用，接触不到诸如微信之类的信息化工具，甚至接触不到网络。信息边缘化是信息化时代一种最为危险的“边缘化”。身处信息化时代，我们的思维方式和语言政策要主动去调整、去顺应。

#### **第四，语言冲突问题。**

语言矛盾容易在两个时期激化，一个是在立国之初，一个是国家经济发展到一定程度开始重视文化的时期。

台湾一个时期一度致力“去中国化”，凡是带中国的东西都要去掉，今天这种情况也没有改变多少。香港一些人，把普通话、简化字上升到意识形态领域，很不理智。汉语国际教育问题仍然是机遇和挑战共存。方言和普通话的矛盾仍然是容易引发地方矛盾的一大问题，特别是在南方地区。西北地区的双语教育问题，如果只把着眼点放在国家通用语言的推广上，可能会出问题。我们国家已经进入语言问题和语言冲突的高发期，要重视可能影响国家统一、民族团结、中国走向世界等方面的语言问题。

此外，语言政策研究应该关注的重点问题有：

1. 语言服务问题；
2. 国际语言生活问题，特别是国际组织和国别的语言政策；
3. 语言教育问题，包括母语教育、华语教育、方言教育、汉语国际传播、外语教育、双语教育、民族语言教育等；
4. 网络空间的语言问题；
5. 社会各领域的语言生活与语言政策，特别是语言与经济、语言与国家安全等；
6. 家庭语言规划问题；
7. 中国语言规划史问题；
8. 话语体系及话语权建设问题；
9. 语言政策和语言规划研究的学科建设问题。

语言政策研究和语言规划研究在中国刚刚开始，过去叫语言文字工作，后来叫语言文字事业，我们才刚刚开始建立学科意识。特别是现在有了四个“皮书”，有了五本专门讨论语言政策和规划的杂志，高校里有了博士点、硕士点来培养这方面的人才。作为语言

政策和语言规划研究人员，我们是行动者、实践者，不应该做空文章，不应该说一些“自我按摩”的话，我们的学术目标，就是为社会语言生活乃至整个社会的发展带来积极影响。

## ► 中心动态

### 语言资源高精尖创新中心召开 学术委员会、咨询委员会会议

2018年1月26日下午，北京语言大学语言资源高精尖创新中心



学术委员会、咨询委员会会议在西郊宾馆召开。教育部原副部长、语言资源高精尖创新中心学术委员会主任赵沁平院士出席会议并主持专家讨论，北京语言大学校长、语言资源高精尖创新中心理事长刘利教授出席会议并讲话，语言资源高精尖创新中心主任李宇明教授出席会议并致辞，中国社会科学院学部委员沈家煊研究员、北京大学陆俭明教授、北京大学俞士汶教授、语言资源高精尖创新中心咨询委员会主任崔希亮教授以及中心其他学术委员、咨询委员出席会议。中心部分项目负责人参加会议。

李宇明主任首先代表中心欢迎各位专家委员拨冗参会，并指出语言资源高精尖创新中心



自成立以来每一步都凝聚着各位专家学者的心血和寄托，在 2017 年中心圆满完成中心理事会、学术委员会、监督委员会、咨询委员会的聘任工作后，中心工作进入了新阶段。2018 年是中心的“提质增效”年，是从 1.0 模式进入 2.0 模式的关键一年，中心将继续营造良好的科研创新环境，加大人才汇聚力度，努力让科学家们获得更高的社会声誉和更多的经济红利。李宇明主任希望与会专家委员能继续为中心发展建言献策，为中国和世界的语言事业贡献智慧和力量。

中心常务副主任杨尔弘教授向各位专家详细汇报了中心 2017 年的工作进展和 2018 年的工作计划，并提出了中心在工作中遇到的困难和问题，请专家们对此进行把脉和指导。她表示中心全体人员将一如既往做好服务工作，与专家学者们共同努力，向着语言资源高精尖创新中心的建设目标稳步前进。



与会专家委员就中心提出的有关议题和面临的难点问题进行了充分深入的研讨，在语言资源采集的规范标准、优先领域、语种顺序等方面提出了深刻见解，同时结合中心规划，为中心探索语言资源学学科建设、进一步做好人才汇聚与培养等工作提出了宝贵的意见和建议。大家一致认为，中心既要服务国家

和北京的战略需要，又要做好学科、研究、人才工作，坚持“顶天立地”的建设理念。



中心学术委员会主任赵沁平院士在总结专家讨论时肯定了中心一年来扎实推进的各项工作，希望中心继续坚持以项目汇人才、以学科留人才的目标，尤其是要进一步加大引进国外顶级人才的力度。他强调中心在不断聚焦任务时，要切实以为北京、为国家、为国际社会服务的原则。赵主任还在语言资源的采集应用、语言资源学学科建设以及人才培养汇聚等方面提出许多新思路、新方向，为中心的相关建设提供了高屋建瓴的指导意见。

语言资源高精尖创新中心领导小组组长、中心理事长刘利校长在总结时首先向各位专家对中心和北语的大力支持表示诚挚的感谢，他充分肯定了中心自成立以来所进行的一系列工作，对中心广泛汇聚人才、推进重大项目、开展国内外学术交流等工作给予了高度称赞。他希望中心能够珍惜这次会议的思想成果，认真梳理，仔细消化，结合“高”“精”“尖”的要求，切实做到工作站位高，实施项目精，产出成果尖，充分发挥中心的创新平台优势，积极推动相关学科、专业、研究工作的深入开展。他表示学校将全力支持中心的发展，力争早日实现中心的建设目标。





## 语言资源高精尖创新中心在日本成功举办 首届“一带一路”语言资源与评测研讨会

2018年5月8日，由北京语言大学语言资源高精尖创新中心（以下简称“中心”）、中国中文信息学会联合组织的首届“一带一路”语言资源与评测研讨会在日本宫崎市成功举办。本次会议是第11届语言资源与评测国际会议（LREC 2018）的一个专题研讨会，共收到来自中国、法国、哈萨克斯坦等国家研究人员提交20余篇论文。



中心主任李宇明教授致辞并做了题为《语言资源的时代需求》主旨报告。他指出，语言资源包括自然语言资源、语言产品资源和语言衍生资源，从语言保护、语言智能、知识挖掘等领域来看，当今社会对语言资源的需求十分迫切，可以说未来是一个语言大数据时代。若要获得语言资源，用好语言资源，需要通过自动标注、自动分析、自主学习等关键技术将语言资源数据化，并建立数据库供各领域专家研究。他指出，语言资源数据化过程势必会带来一些伦理学问题，因此非常有必要组织“未来社会的伦理学论坛”，由各个学科人士共同讨论未来社会的“马龙头、刹车板”等问题，通过发展新的伦理学以保证未来人工智能时代的人类安全。李主任的报告引起与会学者的强烈反响。

会议期间，李宇明主任、杨尔弘教授与欧盟委员会通讯网络与技术司主任盖尔·肯特女士，欧盟委员会通讯网络与技术司多语服

务部负责人马可·马塞拉先生，欧洲语言资源联盟荣誉主席、语言资源与评测国际会议主席尼古丽塔·卡罗拉女士，欧洲语言资源联盟秘书长凯莱德·考克博士进行了座谈，双方各自介绍了目前开展项目的情况，交流了对一带一路沿线国家、欧洲语言资源建设、技术开发的合作意向。会议当天，LREC2018 组委会就在 Twitter 上对本次座谈进行了报道。



本次研讨会除了论文报告，还设计了论坛环节，与会人员就“稀缺语言资源的建设、技术、共享”话题进行了深入交流。杨尔弘教授、综合办公室主任岳岩副教授分别主持了邀请报告和会务工作。

语言资源和评测国际会议在业界影响力很大，每两年举办一次。今年5月7日至12日在日本宫崎召开的 LREC 2018 是第十一届会议，也是该会议首次在亚洲举办。“一带一路”语言资源与评测研讨会是语言资源高精尖创新中心首次在海外主办的国际学术会议，旨在吸引全球相关领域研究人员参与中心建设，探索语言资源采集、加工、保存和共享机制，促进语言资源标准和规范的制定，推动语言工具和技术在语言资源方面的应用。



## 法国著名汉学家白乐桑教授受聘为 语言资源高精尖创新中心国际学术顾问

2018年5月3日上午，语言资源高精尖创新中心（以下简称“中心”）在北京语言大学综合楼810会议室隆重举行仪式聘请白乐桑教授担任中心国际学术顾问。中心主任李宇明教授出席聘任仪式并为白乐桑教授颁发聘书，北京语言大学校长助理、汉语国际教育学部主任张旺喜教授，北京语言大学刘珣教授、崔永华教授、姜丽萍教授参加了聘任仪式。仪式由中心常务副主任杨尔弘教授主持。

李宇明主任在致辞中首先高度赞扬了白乐桑教授在推进法国和欧洲汉语文化教育方面做出的突出贡献，并对白乐桑教授热情接受中心国际学术顾问聘任表示衷心感谢。李主任简要介绍了语言资源高精尖中心成立的



背景、功能定位和未来发展，他指出，作为北京市唯一一个以语言学为背景学科的高精尖中心，中心致力于解决当前学术界关心的语言资源与语言智能等重要问题，重点实施语言资源库、语言文化博物馆和“语言通”智能服务（简称“库”“馆”“通”）三大工



程，搭建平台，汇聚人才，服务以“一带一路”倡议为代表的“人类命运共同体”建设和首都北京“四个中心”建设。李宇明主任表示，衷心希望白乐桑教授

作为“北语的好校友”和“中心的好朋友”，给予中心更多的指导

和支持。随后李宇明主任向白乐桑教授颁发了聘书，并赠送了纪念品。



白乐桑教授非常感谢李宇明主任的邀请以及中心的聘任，他追溯了与中心长久以来的缘分，从慕课教学心得、“幽默与语言”的见解、各国语言教学实际、“第二文字教学”等方面进行了阐述。他表示，特别高兴担任中心国际学术顾问，将利用自己的学术影响继续大力推进中欧汉语国际文化教育，为促进中心与欧洲语言文化学术机构开展交流合作贡献积极力量。

座谈会上，与会专家还围绕全球汉语智能学习、欧洲汉语学习数据的开发利用、慕课与语言保护等主题进行了探讨和交流。

最后，校长助理张旺喜教授作了总结发言。他代表学校和中心领导小组对白乐桑教授受聘为中心国际学术顾问表示衷心祝贺，对白乐桑教授的大力支持表示由衷感谢，希望白乐桑教授给予中心更全面的指导，期待双方未来有更多合作，在助力汉语国际教育、中华文化推广的同时，进一步拓宽中法、中欧友谊的桥梁。



## 澳门理工学院代表团访问语言资源高精尖创新中心

2018年5月5日上午，语言资源高精尖创新中心（以下简称“中心”）在北京语言大学综合楼810会议室接待澳门理工学院院

长李向玉代表团一行。中心主任李宇明教授出席会见仪式并致辞，北京语言大学校长助理、港澳台办公室主任张宝钧教授和中心常务副主任杨尔弘教授参加会见。



李宇明主任对李向玉院长来访表示热烈欢迎。他指出，澳门理工学院在澳门回归、大陆与澳门学术交流合作中做出了积极贡献，北京语言大学与澳门理工学院作为兄弟院校，多年来双方开展了多领域全方位合作，特别是在葡萄牙语教学方面。他还简要介绍了中心基本情况，指出中心致力于实施语言资源库、语言文化博物馆和“语言通”智能服务（简称“库”“馆”“通”）三大工程，并提出“世界知识中文表达”这一新学术理念。李宇明主任表示，感谢李向玉院长一直以来对北语和中心工作的支持，希望未来继续加强与澳门理工学院的交流与合作，更好地推动中心和澳门理工学院的共同发展。

李向玉院长表示，感谢李宇明主任和中心的热情邀请。澳门理工学院在葡萄牙语研究方面有比较深厚的历史积淀，为中国和葡语国家合作交流做出了一定贡献。学院汇聚了国际和中国大陆诸多著名院校的计算机领域人才，开发出一款中葡在线机器翻译软件，目前已服务于澳门特区政府，受到特区政府的高度好评。



双方就中葡英机器翻译联合实验室、中葡语言研究与翻译、人才交流与合作等方面充分交换意见并进行深入探讨。



澳门理工学院社会经济与公共政策所陈庆云教授、语言暨翻译高等学校韩丽丽博士、学术事务部部长李雁莲博士随行来访。北京语言大学港澳台办副主任蔺劼、语言资源高精尖创新中心副主任田列朋参加了会见。

## 针对成人学习者的汉字教学法研究 ——“语言资源与智能”大讲坛第十六讲

2018年3月27日，“语言资源与智能”大讲坛（第十六讲）在北京语言大学综合楼1211会议室开讲。语言资源高精尖创新中心邀请德国哥廷根大学学术孔子学院德方院长、德语区汉语教学协会会长顾安达（Andreas Guder）教授为北语师生作了题为《针对成人学习者的汉字教学法研究》的学术讲座。

顾安达教授认为，欧洲的文字环境都是以字母书写为主，在此影响下，汉字在汉语作为外语教学研究领域内很长时间未得到足够重视，其复杂性往往被低估。他在演讲中分享了采用学习者母语来介绍汉字字符现象的方法，用于提升成人的识字能力。另外，顾安达教授对汉语的“口头语言能力”及“书写语言能力”进行了区分，认为“字符能力”之于书写语言能力就如同“音符能力”之于口头语言能力，而“字符能力”的获取是汉语学习者提高阅读及写

作能力不可或缺的前提。讲座以欧洲汉语能力基准项目（EBCL）为例，以 EBCL 中的“汉语字符能力分级参考”为标准，详细介绍了欧洲汉语学习者为了保持汉语学习动机及学习成效，在汉语学习的不同阶段应掌握的汉字系统知识概况。



顾安达教授的讲座内容充实，研究视角新颖，给在场师生以很大启迪，使大家对欧洲成人汉语教学现状及教学法有了更深刻的理解。讲座后，顾安达教授与在场师生进行互动，交流观点，并对大家提出的问题做了详细解答。

## ► 项目进展

### 《汉学研究》资源建设项目简介

项目负责人 阎纯德教授

为了配合国家“一带一路”、“中国文化走出去”战略方针，加强新时期的中外文化交流，发挥北京语言大学的学科资源和人才优势，以服务国家和首都发展战略为导向，立足北京、辐射全球、放眼全球，打造国际交流与合作的平台，收集国内外有关高校、科研院所、社会团体等研究人员等有关汉学研究的成果，本项目拟以《汉学研究》为依托，汇集海外汉学和海外中国学研究成果。

从13世纪马可·波罗通过陆路往返欧亚起，西方世界便对中国知识、中国形象等有所了解，经历了数代学者的努力耕耘，在文化交流日益频繁的时代语境下，海内外研究者对汉学的兴趣日益浓厚，越来越多的海外学者对中国的哲学、历史学、文学、教育学、政治学、社会学、民俗学等产生兴趣，如宇文所安、梅维恒、艾朗诺等，出版了《哈佛中国史》、《哥伦比亚中国文学史》、《剑桥中国文学史》等书籍，国内更是有阎纯德、李明滨、严绍盪、耿昇、张西平等学者为汉学近三十年来的发展做出了巨大贡献。国内不少大学也设立了汉学研究所或研究中心，培养研究生和博士生，投身于汉学研究的年轻学者日益增多，百余家人文学术期刊开始发表汉学研究的专业论文，《汉学研究》对于汉学家、汉学研究者来说是一个重要的学术阵地。

《汉学研究》创办于1995年，主编为北京语言大学阎纯德教授。阎纯德教授认为，“汉学其本质是外国文化，但是，汉学是以中国文化为原料，经过另外一种文化精神和智慧的加工而形成的一种文化。从某种意义上说，汉学既是外国化了的中国文化，又是中国化了的的外国文化，抑或说是一种亦中亦西、不中不西，有着独立个性的文化。汉学作为一门独立的具有跨文化性质的学科，是外国文化对中国文化研究、阐释、借鉴的结果。”在中国展开汉学研究，不仅可以丰富中国文化，而且在探究中国文化在世界传播发展的路径时还可以赋予其全球性特性。《汉学研究》经历了二十多年的耕耘，现为半年刊，每年出版2集，每集65万字左右，2014年成为CSSCI收录集刊（2014-2015）的来源集刊，现为CSSCI收录集刊（2017-2018）的来源集刊。《汉学研究》现为知名学者魏柳南、



耿昇、张西平开设了专栏，除了介绍美国、法国、澳大利亚等国汉学研究新动态，亦刊发了朝鲜半岛、西班牙、阿拉伯等地汉学研究成果。此外，还设立了汉学家论坛、汉学家研究、汉学视野与汉语教学等专栏，更为全面地反映全球化视野下汉学研究的成果。

北京语言大学素有“汉学家摇篮”之称，在项目实施的具体阶段，《汉学研究》将依托北京语言大学的优势资源，建立以《汉学研究》编辑部为核心的工作团队，并根据实际需要开展中外合作。

## 中华汉字文化体验园项目正式启动

2018年1月20日，北京语言大学语言资源高精尖创新中心“中华汉字文化体验园”项目启动大会在北京国际汉语研修学院隆重召开。语言资源高精尖创新中心主任李宇明教授、北京市教委科研处处长张宪国、国家汉办政策研究处处长王甬等领导出席启动仪式并讲话，来自北京语言大学、清华大学、北京大学等相关领域的专家学者、外部合作单位代表共计50余位嘉宾参加活动。



李宇明主任对项目的成功启动表示热烈祝贺。他指出，语言资源从广义上来说分为三个部分：自然语言资源、语言应用产生的资源、语言知识以及语言技术产生的资源，无论是在工业化时代，还是在信息化时代，语言资源都起着非常重要的作用。语言资源高精尖创新中心对中华汉字文化体验园项目充满信心，同时期望体验园更多关注汉字与科技、汉字与民

族、汉字与国家的关系。语言资源高精尖创新中心是北语的科研中心，更是北京市的高精尖中心，中心将严格按照北京市相关要求做好高精尖项目建设，网罗四海人才，搭建重要平台，为国家和北京市的科学发展做出更大的贡献。

张宪国处长指出，此次启动大会是一场汉字文化的盛宴，中华汉字文化体验园有很好的理念、很好的起点、很好的平台，北京市教委科研处将全力为高精尖项目做好各项服务。

启动仪式上，北京国际汉语研修学院耿学超院长致开幕词，桂帆副院长从项目目标、项目特色和项目预期成果三个方面对体验园进行了简要介绍。此外，来自清华大学美术学院、中国古文字艺术研究中心的陈楠教授，北京大学对外汉语教育学院徐晶凝教授以及葡萄牙 IT 技术专家费尔南多先生分别从汉字文化体验的艺术设计、如何讲好汉字故事、现代展陈技术的应用三个角度，分享对体验园项目的思考与建议。

启动仪式后，项目组专家学者还参加了体验园项目研讨会，从设计规划、内容呈现、科技应用、媒体传播、产品研发、跨界合作等方面



进行了研讨，建设思路逐步清晰，并初步明确独立场馆的建设与项目结构的顶层设计同步进行，确保体验园项目在规划时间内取得实际成效。

中华汉字文化体验园项目作为北京语言大学语言资源高精尖创新中心的重点项目之一，其核心目标是建成一个以宣传推广中国文化为核心的，以汉字汉语教育为基础的，以创新汉字汉语学习模式为特色的寓教于乐综合性体验式学习平台。体验园将以创新汉字教学模式为建园理念，设立汉字历程、汉字创设和汉字文化三大展区。语言资源高精尖创新中心将继续努力做好高精尖项目建设工作，为国家和北京市文化科学事业的发展做出积极贡献。

## 中俄经贸合作数据库及俄汉语智能化综合服务平台 项目顺利完成定期检查

2018年4月21日，语言资源高精尖创新中心（以下简称“中心”）“中俄经贸合作数据库及俄汉语智能化综合服务平台”（以下简称中俄“语商通”）项目进展报告会暨工作研讨会在北京语言大学综合楼909会议室召开。语言资源高精尖创新中心主任李宇明



教授出席会议并致辞，中俄“语商通”项目负责人傅兴尚教授及部分核心成员参加研讨会。中心学术

委员会委员、中国科学院自动化所研究员宗成庆，北京语言大学高级翻译学院韩林涛老师出席会议并与项目组做深入交流。中心副主任田列朋主持会议。

李宇明教授在致辞中指出，语言资源高精尖创新中心作为北京市唯一一个以语言命名的高精尖中心，是为攻克科学难题、汇聚顶

尖人才资源而成立的科研中心。中心主要建设语言资源库、语言文化博物馆和“语言通”智能服务三大工程，即语言资源采集、语言资源展示以及语言资源智能研发利用，简称“库”、“馆”、“通”。

中俄“语商通”项目是“库”和“通”相结合的重要示范，中心以此项目为基础，未来将从俄语扩展到其他语种，从经贸扩展到其他领域，重点研究解决领域语言数据和领域信息数据的汇聚和应用问题。

李主任指出，大数据时代背景下，我们应该提高对数据的认识，数据必须能够有用才能成为资源。近些年人们对“语言是一种资源”的意识较以前增强了，更多人认识到了语言资源的



“用”处。这种“用”主要包含三方面：一是“可用”，二是“能用”，三是“便用”。目前，如何解决非汉语数据的可用和能用的问题是中国人面临的一个大问题。基于此中心近期提出“世界知识的中文表达”理念，希望通过最大努力，让世界知识都能进行中文表达，这既需要技术支撑，也需要国家政策支持，最终目的是希望促进汉语成为国人乃至世界语言的“路径依赖语”。至于“便用”，就是要把信息便捷地推送到有需要的用户手上。过去强调信息的原创，但在信息大爆炸时代，对信息的再加工再推送也是重要劳动。除了生产信息有价值之外，对信息的再加工再处理也成为重

要产业。信息再加工就是把不可用的信息变为可用的信息，把不使用的信息变成使用的信息。



报告会上，中俄“语言通”项目负责人傅兴尚教授结合项目总体构架，介绍了项目的进展情况，并通过其整体构架、资源库建设、平台设计、实施原则、工程保障、前景应用等方面具体解析了“语言通 <=> 领域通”模型涉及的诸多语言学和信息学问题，并指出了该

模型在未来语言智能系统研发、领域知识库建设及二者互促互动中所具有的重要意义。项目组其他成员分别针对中俄“语商通”的框架结构与功能、中俄商贸仿真移动谈判室、中俄语句对齐有关理论与工具、句对库清洗软件等内容的进展进行了详细的介绍与演示。

会上，宗成庆研究员与韩林涛老师针对项目的框架和技术方面等有关问题与项目成员进行了深入的交流，提出很多富有建设性的指导意见。项目组成员围绕专家意见并根据项目进展情况在下午的专题研讨会上进行了针对性的研究和调整。



根据工作安排，4月26日上午，中俄“语言通”项目负责人傅兴尚教授做客“语言资源与智能”大讲坛（第十七讲），为广大师生作了题为《语言通<=>领域通模式解析和实现》的学术讲座，语言资源高精尖创新中心常务副主任杨尔弘教授主持了讲座。

傅兴尚教授认为，语言数据是世界的本源，语言是用来表达世界、认知世界的工具，更是彼此交流的工具。语言通是模拟人的语言能力的智能语言系统，领域通指的是实现信息服务功能的平台，二者结合，充分体现了项目的目标成果。傅兴尚教授还详细介绍了中俄语言通和领域通的内部架构及建设思路，并表示该项目将在助力实施“一带一路”战略的实施和中俄经贸和相互投资的务实合作方面发挥重要作用。

报告会后，中心常务副主任杨尔弘教授和副主任刘晓海副研究员与傅兴尚教授就项目应进一步聚焦语言和信息资源建设、重点解决现实应用需求等方面深入交换了意见。

语言资源高精尖创新中心对下设项目进行定期检查是一项创新性的科研项目管理和服制度，中心各项目根据有关办法和规范要求，每半年举办一次项目工作进展座谈会和多次学术交流会，以学术活动为主要形式保障项目顺利实施，取得了良好的效果。

## **中国周边（6国）语言资源库建设项目工作研讨会 顺利举行**

2018年4月22日，语言资源高精尖创新中心（以下简称“中心”）“中国周边（6国）语言资源库建设”项目工作研讨会在北京语言大学综合楼810会议室召开。中心学术委员会委员李锦芳教授、黄成龙研究员出席会议并与项目负责人郭风岚教授及项目组核心成员进行深入交流。会议由中心副主任刘晓海副研究员主持。

工作会上，项目负责人郭风岚教授介绍了项目进展情况以及项目论证确定实施的四项任务，即（1）目标国三语（国家通用语、地

区通用语、濒危语言)调查; (2) 目标国语言文化调查; (3) 目标国语言文献汇聚;

(4) 目标国语言状况调研。郭风岚教授指出, 该项目具有很强的探索性和较高的实施难度。对此, 各子课题负责人均表示将

攻坚克难按计划完成已开展的工作。



中心学术委员会委员李锦芳教授、黄成龙研究员分别就项目任务设计、调查内容设计及田野调查方法提出很多富有建设性的指导意见。中心副主任刘晓海副研究员对项目组专家的辛勤工作表示感谢, 指出该项目的一个重要意义是将中国语言调查的学术规范标准推向世界, 希望项目组今后继续秉承这一理念, 并预祝项目进展顺利。



语言资源高精尖创新中心根据有关办法和规范要求, 每半年举办一次项目工作进展座谈会和多次学术交流会, 以学术活动为主要形式推进各项目顺利实施, 是一项创新性的科研项目管理和服务制度, 在具体实施中取得了良好的效果。

## 面向智能语音教学的汉语中介语语料库建设 项目进展报告会暨工作研讨会顺利举行

2018年4月26日，语言资源高精尖创新中心（以下简称“中心”）“面向智能语音教学的汉语中介语语料库建设”项目进展报告会暨工作研讨会在北京语言大学主南304会议室召开。中心常务



副主任杨尔弘教授、中心检查组专家姜丽萍教授、中心副主任田列朋出席会议与项目负责人张劲松教授及项目核心成员进行深入交流。会议由中心副主任刘晓海副研

究员主持。

报告会上，项目负责人张劲松教授详细介绍了项目进展、阶段性成果以及成果应用情况。项目进行一年多来，共发表高质量论文二十多篇，其中近十篇发表在国际核心期刊。项目开发的汉语声调教学APP“尔雅中文”已进行为期一个多月的课堂实验，效果良好。项目核心成员在补充发言时介绍了语料库标注情况以及项目组与荷兰莱顿大学、西北师范大学、新疆大学的合作事宜。中心检查组针就项目的具体技术路线和项目成果应用转化共享等问题提出很多富有建设性的意见，对项目下一步的实施提了许多指导建议。

根据工作安排，4月27日上午，“面向智能语音教学的汉语中介语语料库建设”项目负责人张劲松教授，项目核心成员、社科院方强副研究员及新疆大学王玮老师分别以项目成果详细展示、语料库标注及发音建模为主题，做了三场精彩的项目学术报告。



张劲松教授以《汉语智能发音教学研究进展——以“尔雅中文”APP为例》为题，详细介绍了研发汉语智能发音教学应用的初衷，分享了“尔雅中文”APP的设计思路和技术路线，并现场展示了APP的使用场景和效果。方强副研究员



以《面向语音教学的发音建模研究》为题，介绍了由中国社会科学院和天津大学承担的发音建模部分工作的进展情况，指出项目利用EMA等设备采集发音器官的形态数据，建立三维的发音器官模型，并根据学习者的语音信号推测学习者发音器官的形状和位置信息，最后以视频动画的方式反馈给学习者。王玮老师以《汉语中介语语音库构建与标注》为题，介绍了BLCU-SAIT汉语中介语语音语料库的设计、语料收集及标注情况，并就语音库的标注问题与在场的师生进行了交流。

## HSK 动态作文语料库 2.0 版重磅发布

“HSK 动态作文语料库”是母语非汉语的外国人参加高等汉语水平考试（HSK 高等）作文考试的答卷语料库，收集了1992-2005年部分外国考生的作文答卷。语料库1.0版收入语料10740篇，约400万字，于2006年12月上线。2008年7月，经修改补充，语料库1.1版语料总数达到11569篇，共计424万字。

由于旧版语料库开发于 10 多年前，技术已显陈旧，难以保证语料库的正常运转，为更好地服务于国内外的教师、科研人员、汉语学习者，北京语言大学语言资源高精尖创新中心接受语料库管理团队的申请，决定资助 HSK 动态作文语料库 2.0 版的建设与开发。

新版语料库系统开发得到北京语言大学原校长、语言资源高精尖创新中心咨询委员会主任、“HSK 动态作文语料库”课题组负责人崔希亮教授的同意和支持，设计和实施由语言资源高精尖创新中心“汉语中介语语料库创新工程”项目组张宝林教授团队负责，北京语言大学信息科学学院荀恩东教授提供语料库 1.1 版设计数据。

新版语料库保留了 1.1 版的全部语料；检索方式除“一般检索”之外，增加了“特定条件检索”；数据统计增加了可视化图形设计；除检索浏览外，用户还可以对发现的语料录入错误和标注错误进行修改；增加了语料自动下载功能（限 500 条）；增加了“个人工作室”；增加了留言功能。