语言资源高精尖创新中心
Beijing Advanced Innovation Center for Language Resources

中国语言资源保护的理念与实践

The Concept and Practice for Protecting Language Resources of China

李宇明
北京语言大学

“第二届语言资源与智能国际学术研讨会”
（北京，2018年12月16日）

Li Yuming
Beijing Language and Culture University

The 2th International Symposium on Language Resources and Intelligence
(Dec. 16th, Beijing)

语言是一种社会资源

Language Is a Social Resource

中国有100多种语言，汉语的方言灿若星空，汉语文献有数千年积累，汗牛充栋，是世界上语言资源十分丰富的国度。

There are over 100 languages and plentiful dialects in China and Chinese literature has been accumulated for thousands of years, which made China a country with rich language resources in the world.

中国在历史上也比较重视对语言资源的保护和开发利用，有许多经验值得总结和继承。

China has been attaching importance to the protection, development and application of language resources throughout history, resulting in many experiences for summarization and inheritance.

语言是一种社会资源

Language Is a Social Resource

语言是一种社会资源

Language Is a Social Resource

中华人民共和国教育部
Ministry of Education of the People's Republic of China

信息公开 | 信息公开规定 | 信息公开指南 | 信息公开目录 | 依申请公开 | 信息公开年度报告

政务之窗 | 机构设置 | 新闻发布 | 公报公告 | 统计数据 | 政府采购 | 专题专栏 | 信 息 化 | 人事任免 | 政策法规 | 文献资料

服务大厅 | 行政审批 | 办事公开 | 就业指导 | 名单查阅 | 学历查询 | 学历认证 | 学位查询 | 学位认证

互动平台 | 部长信箱 | 政策咨询 | 专家答疑 | 政策解读 | 征求意见 | 在线访谈 | 热线电话 | 滇西开发 | 移动客户端 | 新闻办微博 微信

【浏览字体： 大 中 小】

教育部 国家语委关于启动中国语言资源保护工程的通知

教语信[2015]2号

各省、自治区、直辖市教育厅（教委）、语委，新疆生产建设兵团教育局、语委，有关省、自治区民委（民语委），有关高校、科研院所：

为贯彻党的十八大和十七届六中全会关于大力推广和规范使用国家通用语言文字，科学保护各民族语言文字的精神，落实《国家中长期语言文字事业改革和发展规划纲要（2012－2020年）》的任务要求，教育部、国家语委决定自2015年起启动中国语言资源保护工程（以下简称语保工程），在全国范围开展以语言资源调查、保存、展示和开发利用等为核心的各项工作。

一、实施基础

为更好地掌握语言国情，保护国家语言资源，传承和弘扬中华优秀传统文化，为国家建设和发展战略提供服务，教育部、国家语委从2008年起，先后在江苏、上海、北京、广西、辽宁、福建、山东、河北、湖北等省份开展了中国语言资源有声数据库建设试点工作。目前，江苏、北京和上海已完成本地区相关语言资源的调查、采录和整理并通过验收，其余相关省份的调查工作正在有序推进。上述省份富有成效的试点工作检验了有声数据库建设技术规范和工作规范的科学性和可行性，完善了中国语言资源有声数据库建设方案，探索出一套“政府主导、学者支持、社会参与”的工作模式和一系列行之有效的专家团队运作及项目管理办法，为在更大范围内开展语言资源保护工作积累了宝贵经验。同时，“中国语言资源有声数据库技术规范与平台研发”项目得到科技部2014年度国家科技支撑计划的支持，为语言资源保护工作提供了有力的技术保障。

党的十八大和十七届六中全会对语言文字工作提出了明确要求，赋予了科学保护各民族语言文字的重大使命。《语言文字规划纲要》将科学保护各民族语言文字列为重要任务。鉴于当前工作任务的需要和前期良好的

中国语言资源研究三阶段

Three Stages of Study on Language Resources of China

01

20世纪80-90年代，认识语言的资源性质。

1980s-1990s Resource attribute of language was cognized.

02

21世纪00-10年代，语言资源的分类及外延。

2000s-2010s Classification and denotation was given to language resources.

03

当下，从功用的视角研究语言资源，关注：语言资源的功用及其发挥，从“用”的视角来评定语言资源的收集、整理、贮存、保护及开发等。

Nowadays, language resource study focuses on its function and realization of the function. It is from the perspective of application to evaluate the language resources collection, organization, curation, protection and development.

语言资源的类型

Types of Language Resources

口头语言资源

（最为基础的语言资源）

Spoken language resources

(the most fundamental language resources)

书面语言资源

（社会作用最大的语言资源）

Written language resources

(language resources playing the most important role in society)

语言衍生资源

语言知识、语言能力、人工语言智能、语言技术、语言艺术等

Derived language resource :

language knowledge, language competence, language intelligence, language technologies, language arts, etc.

语言衍生资源

Derived Language Resources

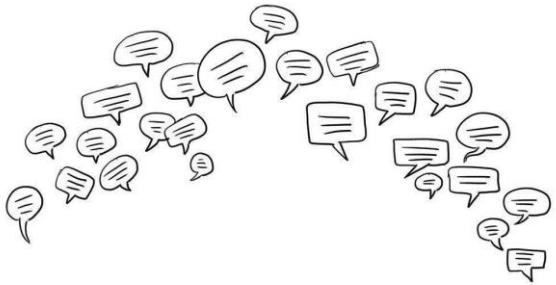
- **语言知识：语言研究的成果**
  Language knowledge: outputs from language study
- **语言能力：人类语言学习的内化**
  Language competence: internalization of human language learning
- **语言智能：机器学习语言的内化**
  Language intelligence: internalization of machine language learning
- **语言技术：辅助语言发挥功能的各种技术**
  Language technologies: various technologies assisting a language to perform its function
- **语言艺术：语言应用于艺术的成果**
  Language arts: outputs from application of language to the field of arts
- ……

语言资源的认识

Cognization of Language Resources

语言资源的认识，是逐步发展的。

Cognization of language resources has been gradually enhanced.

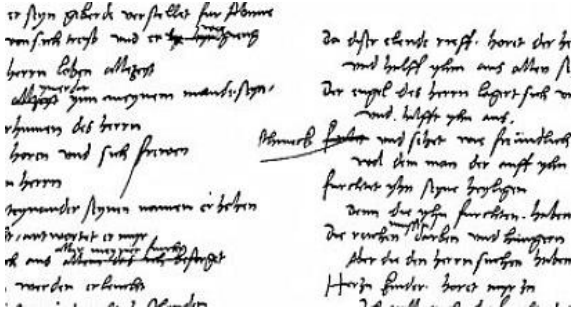


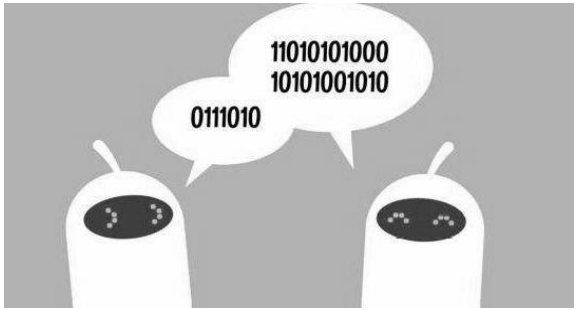
首先认识的是口头语言资源

Spoken language resources were firstly cognized.

其次认识的是书面语言资源

Written language resources were later cognized.





语言衍生资源，现在还在被认识之中

Derived language resources are still under cognization.

自然语言是个“知识系统”

Natural Language is a Knowledge System

- 自然语言（口语和书面语），能够作为资源，能够充分发挥资源的作用，不仅是个“符号系统”，更是个“知识系统”。
- Natural language (spoken and written) can be regarded as resources and fully perform the resource function, so it is a “knowledge system” more than a “semiotic system”.

- 语言保护不只是记录、保护语言符号，而更要记录、保护人类的语言知识系统。
- Language protection requires not only recording and protecting language sign, but recording and protecting human language knowledge system.

自然语言是个“知识系统”

Natural Language is a Knowledge System

- 传统的语言调查，目的是揭示语言的符号系统，而相对忽视调查研究语言的知识体系。
- Traditional language investigation is to reveal language semiotic systems, but ignores studying its knowledge system.

- “语言知识观”，有助于认识语言的资源属性，会有崭新的语言调查深度，比如：对于词汇的调查，就不仅仅是在常用词的范围，对于语法、语用的调查，就会更深入地涉及到文化与心理。
- The knowledge view of language benefits the cognization of the resource attribute of language, and deepens language investigation. So the examination of vocabulary will not be restricted in the scope of commonly used words, and inquiry into the grammar and pragmatics will involve more cultural and psychological elements.

语言保护三层次

Three Levels of Language Protection

01

“语言保存”或曰“语言资料保存”

Language preservation or language data preservation

02

“语言保育”或曰“语言生态保育”

Language conservation or language ecosystem conservation

03

语言资源开发

Language resources development

语言保护三层次

Three Levels of Language Protection

“语言保存”或“语言资料保存”
Language preservation or language data preservation

通过书面方式和录音录像方式，将语言记录下来，建库保存下来。当前学者进行的多是语言保存层面的工作。

record and preserve the language in written form and in recordings and videos. What most scholars do today is in the level of language preservation.

“语言保育”或“语言生态保育”
Language conservation or language ecosystem conservation

通过各种措施来延长语言生命、维护语言活力、保持语言的生态环境。

Language conservation or language ecosystem conservation: Take various measures to keep language alive, and maintain language vitality and ecological environment.

语言资源开发
Language resources development

对语言保存、语言保育成果的进一步开发利用，获取语言保护的“红利”。

Language resources development: Further develop and apply the achievements of language preservation and conservation to obtain the bonus of language protection.

指向语言智能的语言资源

Language Resources for Language Intelligence

01

社会正在向“智能时代”迈进，人工智能是智能时代最主要的技术力量。
Our society is towards the era of intelligence. Artificial intelligence is the prime technological power in the era of intelligence.

02

人工智能的核心是语言智能，机器获取语言智能主要靠语言大数据的训练。
The core of artificial intelligence is language intelligence. Machine acquires language intelligence mainly by big language date training.

03

语言大数据也是语言资源，从语言智能的视角来观照语言资源及其保护，语言资源就进入了生产资料的范畴，对人类的意义就更加不一般了。
Big language data is also language resources. If we scan language resources and the protection from the view of language intelligence, language resources can be grouped in the category of means of production with even more special meaning to humans.

语言资源的共建共享

Co-contruction and Sharing of Language Resources

语言资源的保护和开发利用，需要人类社会的合作，包括不同地区、不同国家、不同国际组织的合作，不同社会部门和不同学科的合作。

The protection, development and application of language resources require cooperation among human communities, including cooperation from different regions, countries and international organizations, and cooperation across different social departments and disciplines.

为保证这种合作的开展，需要制定一系列国际标准，包括技术标准、工作标准。

To ensure the cooperation, a series of international standards (technology standard, work standard, etc.) will be made.

2018年9月19日，由联合国教科文组织、中国教育部、湖南省人民政府等联合主办的首届世界语言资源保护大会在湖南长沙开幕。

International Conference “Role of Linguistic Diversity in Building a Global Community with Shared Future” was held on September 19th, 2018 in Changsha, Hunan Province. The Conference was co-organized by UNESCO, Ministry of Education of China and People’s Government of Hunan Province.

世界语言资源保护大会形成了重要成果《岳麓宣言（草案）》，这是联合国教科文组织首个以“保护语言多样性”为主题的宣言。

Yuelu Declaration (Draft) was resulted as an important achievement of the Conference. It is the first Declaration with the theme of Protecting Linguistic Diversity” in UNESCO.

语言资源高精尖创新中心
Beijing Advanced Innovation Center for Language Resources

谢 谢 大 家

Thank you for listening