

语言资源构建

理论、方法与应用国际研讨会

International Workshop on Language Resource Construction

Theory, Methodology and Applications

会议地点：中国北京

会议时间：2017年11月5日、6日

November 5 — 6, 2017, Beijing, China

主办单位

Organizers

计算语言学教育部重点实验室（北京大学）

Key Laboratory of Computational Linguistics (Peking University)

Ministry of Education

北京大学中文系 中国语言学研究

Department of Chinese Language and Literature PKU

Center for Chinese Linguistics PKU

美国宾州大学语言资源联盟（LDC）

Linguistic Data Consortium, University of Pennsylvania

前言

在自然语言信息处理旺盛应用需求的推动下，自然语言知识资源的建设历经 30 多年的高速发展，已经积累了相当丰富的数据。如能对语言资源构建的理论及方法加以总结和提炼，可以更好地指导今后的工作。而在当前深度学习的热潮下，如何认识语言资源数据加工的意义和价值，如何更有效地组织语言资源建设，都是非常值得探讨的议题。从实际应用的角度，从多学科发展的不同视角来审视自然语言知识资源建设的现状和未来发展，无疑是极具有前瞻性和挑战性的论题。

计算语言学教育部重点实验室（北京大学）、北京大学中文系、中国语言学研究中国与美国宾州大学语言资源联盟（LDC）联合召开本次会议，邀请相关专业背景的专家学者，就语言资源建设的理论、方法及应用前景展开深度交流，共同推动自然语言知识资源的未来发展。

会议将采取主题发言和讨论的形式对以下议题展开探讨：（1）语言资源构建的理论、方法与技术；（2）语言资源与深度学习；（3）语言资源在语言计算与社会服务中的应用。并将于会议期间举行相关论坛，探讨在语言资源方面深度合作的内容与方式。

Preface

Driven by the strong demand for applications of natural language information processing, the construction of natural language and knowledge resources has rapidly developed over the past 30 years and accumulated a wealth of data. It is the time now to consolidate and synthesize the existing theories and methods of language resource construction, providing new guidance for future research. And the current boom of “deep learning” methods sparks a debate about the meaning and value of traditional language resource data processing and about the opportunities and challenges to construct language resources more efficiently. To address these issues, the Key Laboratory of Computational Linguistics(Peking University), Department of Chinese Language & Literature , Center for Chinese Linguistics, Peking University, and Linguistic Data Consortium, University of Pennsylvania have co — organized this workshop, which invites experts in related research areas to present results and proposals, and to promote the future development of language resource construction.

The following topics will be discussed in the form of talks and round — table discussions: 1. Theory, methodology and applications of language resource construction; 2. Language resources and deep learning; 3. Applications of language resources in technological, scientific, educational, and clinical areas. The workshop will also provide forums for participants to network and discuss collaborations.

会议议程

Conference Agenda

Day1: 4/11,2017

14:00 — 24:00	外地代表入住(Check In)
----------------------	------------------

Day2: 5/11,2017

北京大学 李兆基人文学苑 1 号楼 108 会议室

Venue: Lee Shau-Kee Academy of Humanities , Room 108 , Building No.1, Peking University

8:00 — 8:30	注册报到 (Registration)
8:30 — 9:00	主管领导致辞 (北京大学科研部、北京大学信息科学技术学院、计算语言学教育部重点实验室、北京大学中文系、中国语言学研究中心、美国宾州大学语言资源联盟 LDC) (Welcoming speech from the Scientific Research Department of Peking University, School of Electronic Engineering and Computer Science PKU , Key Laboratory of Computational Linguistics (Peking University), Department of Chinese Language & Literature , Center for Chinese Linguistics, Peking University, and Linguistic Data Consortium, University of Pennsylvania)
9:00 — 9:20	合影(Group Photo)

<p>9:20 — 10:40</p> <p>每人限时 20 分钟 (20 minutes for each)</p>	<p>演讲 (主持人 : 李宇明教授)</p> <p>Keynote Speeches Session Chair: Li Yuming</p> <p>演讲 1 :</p> <p>语言是一种复杂系统 : 基于认识语言行为机制的资源构建 (陈保亚 Chen Baoya) Language is a complex system: the resource construction based on language cognitive behavior</p> <p>演讲 2 :</p> <p>中文深层语义描述及资源建设 (穗志方 Sui Zhifang) Chinese Deep Semantic Representation and the Resource Construction</p> <p>演讲 3 :</p> <p>跨语言知识图谱构造 (李涓子 Li Juanzi) Cross Lingual Knowledge Graph Building</p> <p>演讲 4 :</p> <p>动词语义标注: 框架为本, 构式为用 (刘美君 Meichun Liu) Semantic annotation: a frame — based constructional approach</p>
<p>10:40 — 10:50</p>	<p>茶歇(Tea Break)</p>
<p>10:50 — 12:10</p> <p>每人限时 20 分钟 (20 minutes for each)</p>	<p>演讲 (主持人 : 孙茂松教授)</p> <p>Keynote Speeches Session Chair: Sun Maosong</p> <p>演讲 1 :</p> <p>标注语料与认知脑神经实验资料整合芻议 (黄居仁 Chu-Ren Huang) Aggregation of Linguistic Annotation and Neuro — Cognitive Data: A Proposal</p> <p>演讲 2 :</p> <p>语言资源建设中激励机制和工作流程的创新 (Chris Cieri) Novel Incentives and Engineering Unique Workflows</p>

	<p>演讲 3 :</p> <p>语义知识资源建设及其在内容计算中的应用 (袁毓林 Yuan Yulin) A semantic knowledge base construction and its content — based computing</p> <p>演讲 4 :</p> <p>面向应用任务的语料库建设 (宗成庆 Zong Chengqing) Task — Oriented Corpus Construction</p>
12:20 — 13:30	<p>午餐(Lunch)</p> <p>地址：北京大学勺园西餐厅地下一层自助厅 Address: Buffet in Shao Yuan Western Restaurant B1, Peking University</p>
<p>14:00 — 15:20</p> <p>每人限时 20 分钟 (20 minutes for each)</p>	<p>演讲 (主持人：宗成庆教授)</p> <p>Keynote Speeches Session Chair:Zong Chengqing</p> <p>演讲 1 :</p> <p>深度学习与自动作诗 (孙茂松 Sun Maosong)</p> <p>演讲 2 :</p> <p>从知识图谱到事理图谱 (刘挺 LiuTing) From Knowledge Graph to Event Evolutionary Graph</p> <p>演讲 3 :</p> <p>面向低资源语言的神经机器翻译 (刘洋 Liu Yang) Neural Machine Translation for Low — Resource Languages</p> <p>演讲 4 :</p> <p>文本自动生成：资源、模型与挑战 (万小军 Wan Xiaojun)</p>
15:20 — 15:30	<p>茶歇(Tea Break)</p>

<p>15:30 — 16:50</p> <p>每人限时 20 分钟 (20 minutes for each)</p>	<p>演讲 (主持人 : 王厚峰教授)</p> <p>Keynote Speeches Session Chair:Wang Houfeng</p> <p>演讲 1 :</p> <p>汉语依存句法数据标注规范、方法、和平台探索 (张民 Zhang Min、李正华 Li Zhenghua)</p> <p>Work Progress on Chinese Dependency Treebanking: Annotation Guideline, Method, and Platform</p> <p>演讲 2 :</p> <p>面向 NLP 的汉语框架网工程 (李茹 Li Ru)</p> <p>Chinese FrameNet Project for NLP</p> <p>演讲 3 :</p> <p>《大词林》中实体类型获取及层次化构建方法 (秦兵 Qinbing)</p> <p>Hypernym of Entity acquire and hierarchy construction in BigCilin</p> <p>演讲 4 :</p> <p>融合概念对齐信息的中文 AMR 语料库的构建 (李斌 Li Bin、曲维光 Qu Weiguang)</p> <p>Construction of Chinese Abstract Meaning Representation Corpus with Concept-to-word Alignment</p>
<p>16:50 — 17:00</p>	<p>茶歇(Tea Break)</p>
<p>17:00 — 18:20</p>	<p>圆桌会议 (主持人 : 詹卫东、常宝宝)</p> <p>(Round-Table Discussion)</p> <p>Session Chair: Zhan Weidong, Chang Baobao</p> <p>“深度学习下的语言资源建设”</p> <p>Language Resources Construction with Deep Learning</p>
<p>18:30 — 20:00</p>	<p>晚餐(Supper)</p> <p>地址 : 北京大学勺园正大国际交流中心中餐厅燕园厅</p>

Address: Shao Yuan Zhengda International Communication Center , Chinese Restaurant ,Yanyuan Hall ,Peking University.

Day3: 6/11,2017

地点：北京宾大沃顿中国中心

Venue: Penn Wharton China Center(Beijing)

地址：北京市朝阳区东三环中路1号北京环球金融中心西塔16层

Address: World Financial Center, West Building, 16F, No. 1 East Third Ring Middle Road , Chaoyang District, Beijing,100020, China.

9:00 — 10:20

**每人限时 20 分钟
(20 minutes for
each)**

演讲 (主持人：黄居仁教授)

Keynote Speeches Session Chair: Chu-Ren Huang

演讲 1：

中国社会科学院语言所语音库建设及相关研究 (李爱军 Li Aijun)
Spoken Language Resources and Phonetic Research at CASS

演讲 2：

多语种平行语音库建设及跨语言韵律研究 (丁红卫 Ding Hongwei)
Multilingual Speech Database Building and Cross-linguistic Prosodic Research

演讲 3：

《中国语言志 普通话》的调研与编写思路 (曹文 Cao Wen)
Survey of Local Putonghua: A Proposal

演讲 4：

大规模语料库语音学研究 (袁家宏 Yuan Jiahong)
Using Large Speech Corpora for Phonetic Research

<p>10:20 — 10:30</p>	<p>茶歇(Tea Break)</p>
<p>10:30 — 12:10 每人限时 20 分钟 (20 minutes for each)</p>	<p>演讲 (主持人 : 李爱军 研究员)</p> <p>Keynote Speeches Session Chair: Li Aijun</p> <p>演讲 1 :</p> <p>粤语病理语言的自动评估与分析 (李丹 Tan Lee) Acoustical Analysis and Automatic Assessment of Pathological Speech of Cantonese</p> <p>演讲 2 :</p> <p>藏语言语数据库的建设及其基于深度学习的研究 (王龙标 Wang Longbiao) Construction of Tibetan Spoken Language Database and Its Study Based on Deep Learning</p> <p>演讲 3 :</p> <p>深度学习在语音信号处理的应用 (杜俊 Du Jun) Deep Learning for Speech Signal Processing</p> <p>演讲 4 :</p> <p>智能汉语发音教学研究(张劲松 Zhang Jinsong) Study on Intelligent Technology for Chinese Pronunciation Teaching</p> <p>演讲 5 :</p> <p>汉语语块获取与应用 (荀恩东 Xun EnDong)</p>
<p>12:10 — 13:10</p>	<p>午餐 (宾大沃顿中国中心自助 (Buffet in Penn Wharton China Center)</p>
<p>14:00 — 15:40 每人限时 20 分钟 (20 minutes for each)</p>	<p>演讲 (主持人 : 陈向京 教授)</p> <p>Keynote Speeches Session Chair: Chen Xiangjing</p> <p>演讲 1 :</p> <p>语料的规模与质量——教育教材语料库建设的思考与启示 (苏新春</p>

	<p>Su Xinchun)</p> <p>演讲 2 :</p> <p>语料库在语言认知研究中的应用 (周晓林 Zhou Xiaolin)</p> <p>演讲 3 :</p> <p>寻找汉语世界的语义基元 (刘扬 Liu Yang) Towards a Description of Chinese Semantic Primitives for Understanding and Computing</p> <p>演讲 4 :</p> <p>现代汉语虚词用法知识库及其应用研究 (咎红英 Zan Hongying) Research on the Modern Chinese Function Word Usage Knowledge Base and it's Applications</p> <p>演讲 5:</p> <p>机器学习算法在翻译风格研究中的应用 (蒋跃 Jiang Yue , 詹菊红 Zan Juhong) Using Machine Learning Methods to Discriminate Translation Styles</p>
15:40 — 15:50	茶歇(Tea Break)
<p>15:50 — 17:10</p> <p>每人限时 20 分钟 (20 minutes for each)</p>	<p>演讲 (主持人 : 袁家宏博士)</p> <p>Keynote Speeches Session Chair: Jiahong Yuan</p> <p>演讲 1 :</p> <p>汉语语保工程和方言文化开发应用 (曹志耘 Cao Zhiyun) Project for the Protection of Language Resources of China and Dialect-culture Application</p> <p>演讲 2 :</p> <p>语言资源高精尖创新中心资源建设 (杨尔弘 Yang Erhong) An Overview of Resources Construction in Beijing Advanced Innovation Center for Language Resources</p> <p>演讲 3 :</p> <p>美国宾州大学语言资源联盟简介 (Denise DiPersio)</p>

	<p>Introduction to Linguistic Data Consortium</p> <p>演讲 4 :</p> <p>人类语言科学与技术的机遇与挑战 (Mark Liberman)</p> <p>Challenges and Opportunities in Human Language Science and Technology</p>
17:20 — 18:20	<p>圆桌会议 (主持人 : 穗志方、 Mark Liberman)</p> <p>(Round-Table Discussion)</p> <p>Session Chair: Sui Zhifang , Mark Liberman</p> <p>“语言资源——合作与共赢”</p> <p>Language Resources:Cooperation and Win-Win</p>
18:30 — 20:00	<p>晚餐 Supper</p> <p>汉舍中国菜馆 Madam Zhu’s Kitchen</p> <p>地址 : 北京市朝阳区建国门外大街 1 号国贸 3 期 B 座北区 L4 层 NL4019</p> <p>NL4019,L4, North Block B, 3 Phase of China World Trade Center,Jianguomen WaiDajie,Chaoyang District,Beijing</p>
20 : 10	<p>搭乘校车返回北京大学</p> <p>(Back To Peking University by school bus)</p>

重要提示(P.S.) :

2017 年 11 月 6 日 (星期一) 统一搭乘北京大学校车前往北京宾大沃顿中国中心。

On Nov.6th(Monday),2017 , We will take the participants to Penn Wharton China Center (Beijing) by Peking University school bus.

集合时间 : 上午 07:30

地点:北京大学英杰交流中心(靠近北京大学东南门)

Time: 07:30 A.M.

Location: Ying Jie Communication Center, Peking University (South-East Gate of Peking University nearby)

会议组织联络人 (Liaisons) :

穗志方 Sui Zhifang (13522038062)

詹卫东 Zhan weidong (13552331912)

姚颖佩 Yao Yingpei (15801675635)

饶琪 Rao Qi (18607110419)

汪梦翔 Wang Mengxiang (13331138412)

王佳俊 Wang Jiajun (13162080618)

会议示意图 Map

