
语言工程与计算实验室

工作简报

(2017 年第 2 期 总第 8 期)

2017 年 12 月 31 日

本期目录

一、实验室要闻

1. 实验室 2017 年喜获国家社科及省部级课题立项
2. 雅典大学计算语言学专家访问实验室开展学术交流活动
3. 语言工程与计算实验室组织 2018 年度高层次项目申报指导交流会
4. 实验室 2017 年招标课题成功立项 3 项课题

二、学术交流

5. 雅典大学专家讲述计算语言学前沿系列讲座
6. 云山讲座教授宗成庆做“面向机器翻译的语料获取方法”的精彩讲座
7. 实验室研究人员参加第 21 届亚洲自然语言处理国际会议
8. 实验室研究人员参加第十六届全国计算语言学会议
9. 实验室研究人员参加第六届国际自然语言处理与中文计算会议
10. 实验室研究人员参加第三届“面向翻译的术语研究”国际学术研讨会
11. 实验室研究人员参加第十三届国际信息检索会议

三、科研成果

12. 实验室 2017 年科学研究汇总

实验室要闻

1. 实验室 2017 年喜获国家社科及省部级课题立项

2017 年 11 月，实验室喜获一项国家社科和两项国家语委/名词委项目，分别是：

国家社科项目：汉语篇章结构的特征—依存描写机制及资源建设研究

国家名词委项目：多维视角下的术语集约性研究

国家语委项目：中国学习者英语花园幽径句的计算语言学研究

实验室致力于学术研究的深入化，充分调动科研人员在深化学术成果方面的积极性，学术研究方向与实验室特色也结合得愈加紧密，此次实验室获国家语委/名词委项目成功立项与相关科研人员的辛苦努力密不可分，期待实验室在深化学术研究、完善研究成果方面取得更大的成绩。

2. 雅典大学计算语言学专家访问实验室开展学术交流活动

2017 年 9 月 26 日上午，雅典大学 Christina Alexandris 副教授参观访问了语言工程与计算实验室并与实验室人员进行了充分地沟通交流。语言工程与计算实验室主任李心广教授、副主任李霞教授及实验室全体研究人员参加了此次交流活动。

李心广主任对 Christina Alexandris 副教授的来访表示了衷心的期盼和热烈的欢迎。他向 Christina Alexandris 副教授介绍了语言工程与计算实验室目前的人员现状、研究方向、取得的研究成果以及实验室目前在研的项目等信息，并表示希望能与雅典大学建立长久的合作关系。Christina Alexandris 副教授详细介绍了雅典大学计算语言学的发展现状和研究方向，重点介绍了其所在实验室“UoA Lab for Computational Linguistics Research”构建的 SKYPE 自动识别和翻译系统。双方就可能合作的研究领域展开了充分的讨论，希望能够通过合著、合编、合写等方式深化计算语言学方面的研究。

3. 语言工程与计算实验室组织 2018 年度高层次项目申报指导交流会

10 月 27 日上午，语言工程与计算实验室在北校区六教 B415 举办 2018 年高层次项目申报交流会，中国科学院自动化研究所模式识别国家重点实验室宗成庆研究员应邀对实验室拟申报的高层次项目进行指导。实验室李心广主任、李霞副主任、实验室全体专职研究员等出席会议。会议由李霞副主任主持。

指导会过程中，首先由每位项目申报者对自己的项目选题、研究内容、研究难点、计划和路线等进行汇报。汇报完毕，宗老师对每位项目申请人的汇报内容和项目申请书进行了专门的指导和针对性的点评。现场交流气氛热烈，各位项目申报者表示专家的针对性指导能够帮助自己理清思路，凝练研究要点，明确研究方向。

4. 实验室 2017 年招标课题成功立项 3 项课题

根据《广东外语外贸大学语言工程与计算实验室招标课题管理实施办法》和《广东外语外贸大学语言工程与计算实验室 2017 年招标课题指南》的有关要求，经申请人自主申报，语言工程与计算实验室初审，实验室学术委员会专家函评、专家会评，实验室主任批准，语言工程与计算实验室 2017 年立项 3 项招标课题。

序号	申请人	课题名称
1	彭三城	公共事件的负面情绪识别及演化分析研究
2	夏立新	基于语料库的中国英语语法本土化研究
3	张新猛	基于网络表示学习的社区主题演化分析

学术交流

5. 雅典大学专家讲述计算语言学前沿系列讲座

2017年9月25日至9月29日，由国际合作与交流处主办，语言工程与计算实验室承办的海外专家短期访学项目“基于欧盟术语框架模式的计算语言学研究”在广东外语外贸大学南校区顺利实施。雅典大学 Christina Alexandris 副教授为广外师生带来了三场精彩的系列学术讲座。讲座活动由语言工程与计算实验室主任李心广教授主持。

9月26日上午，Christina Alexandris 副教授开展了系列讲座中的第一场讲座。讲座的主题是：复合词和词簇术语研究：以欧洲语言为例。她以复合词和词簇为例讨论了词头在尾的德语文本术语以及词头在首的法语文本术语的形成规律，并实践分析了机器处理不同词头文本过程中所面临的问题，并提出了解决方案。

9月27日上午，Christina Alexandris 副教授讲述了基于欧盟多语言系统的术语审定的语言和实践研究。她具体分析了欧盟成员国语言的多样性，包括各自所属的语系、语族和书写系统之间的内在联系，并重点分析了欧盟在多语言环境下如何实现术语的统一和审定，讨论了术语的稳定性和变异性。Christina Alexandris 副教授还现场演示了欧盟 IATE 术语数据库的在线检索和多语种查询。

9月27日下午，Christina Alexandris 副教授做了题为新闻转录文本的机器翻译研究的最后一场讲座。讲座中，她从语法、语义和语用层面分析了新闻转录文本中的副语言信息，讨论了新闻发言中韵律的重要性，并以多个信息检索工具对转录文本进行了语义标记，并对歧义消解进行了实例分析。

6. 云山讲座教授宗成庆做“面向机器翻译的语料获取方法”的精彩讲座

2017年10月27日上午，在广东外语外贸大学北校区六教 B415，宗教授做了一场题为“面向机器翻译的语料获取方法”的学术报告。实验室李心广主任、李霞副主任、实验室全体人员以及学校各相关学院（研究中心）40余位师生聆听了本次报告。本次报告由语言工程与计算实验室主任李心广主持。

宗教授首先指出统计机器翻译（SMT）和神经机器翻译（NMT）的局限性。SMT 和 NMT 近年来取得了长足的进步，但这两种翻译方法（尤其是 NMT）需要大规模双语平行语料，而对于很多小语种之间的翻译来说通常很难得到大规模平行句对，因此，如何利用互联网数据自动获取双语平行句对或翻译片段，成为目前数据驱动的机器翻译方法研究的一种“瓶颈技术”。随后，宗教授在报告上介绍了 SMT 和 NMT 基本方法及已有数据资源的基础上，介绍一种平行句对和翻

译片段两级平行资源同时获取的方法，并对机器翻译目前存在的问题进行分析。最后，他重点介绍了他们团队目前开发的一个多模态语料库以及取得相关成果。

7. 实验室研究人员参加第 21 届亚洲自然语言处理国际会议

2017 年 12 月 4 日-6 日，实验室教师李霞教授赴新加坡参加第 21 届亚洲自然语言处理国际会议 IALP2017。该会议专注于亚洲自然语言处理方面的技术和理论，会议第一届于 1986 年在新加坡召开，截止 2017 年已经举办了 21 届。会议主题包括多模态表示和处理、语义分析、分词、句法分析、篇章分析、文本分析、机器翻译等主题。李霞教授的论文“Chinese-Thai Cross Language Event Extraction and Alignment”被会议录用，并在会议中做了论文报告。同时李霞教授还主持了会议的第 10 个 Session，主题为 Text Mining，该 Session 的论文包括来自中国、日本和新加坡等国家的论文汇报。

8. 实验室研究人员参加第十六届全国计算语言学会议

“第十六届全国计算语言学学术会议”（The Sixteenth China National Conference on Computational Linguistics, CCL 2017）于 2017 年 10 月 13 日—15 日在江苏南京举行。作为国内最大的自然语言处理专家学者的社团组织——中国中文信息学会（CIPS）的旗舰会议，CCL 着重于中国境内各类语言的计算处理，为研讨和传播计算语言学最新的学术和技术成果提供了高水平的深入交流平台。

实验室研究人员李心广教授、李霞教授、云山青年学者任函博士、冯文贺博士和兼职研究员郝天永副教授、丘心颖副教授参加了此次会议。另外，实验室学术委员会主任孙茂松教授、学术委员会委员刘挺教授和周国栋教授也参加了此次会议。会议期间，实验室李心广主任、李霞副主任拜访了实验室学术委员会委员，就实验室 2017 年实验室招标课题进行了讨论。

此次会议实验室李霞、任函和冯文贺三位老师的四篇论文被录用，具体信息如下：

1. 李霞. 基于局部密度的作文跑题检测方法. CCL2017
2. 任函. Natural Logic Inference for Emotion Detection. CCL2017
3. 冯文贺. 《文心雕龙》的篇章连接词. CCL2017
4. 冯文贺. 汉语小句的俄语对应单位研究. CCL2017

9. 实验室研究人员参加第六届国际自然语言处理与中文计算会议

第六届 CCF 国际自然语言处理与中文计算会议（NLPC2017）于 2017 年 11 月 8 日-11 月 12 日在大连举行。该会是由中国计算机学会（CCF）主办的 CCF 中文信息技术专业委员会年度学术会议，是专注于自然语言处理及中文计算领域的国际会议。本次大会关注语言计算领域的研究和创新成果，邀请国内外知名学者分享特邀报告和相关领域前沿讲习班课程，并安排学术论文报告、技术测评报告、Poster 和工业展览及报告等。

实验室李心广教授、李霞教授，兼职研究员蒋盛益教授、周咏梅教授、柯晓华教授、郝天永副教授、丘心颖副教授参加了会议。实验室学术委员会委员宗成庆教授、周国栋教授参加了会议。

此次会议实验室李霞老师的论文被录用并做了论文宣讲，具体信息如下：

Xia Li et al. Chinese Question Classification Based on Joint Semantic Features. NLPCC2017.

10. 实验室研究人员参加第三届“面向翻译的术语研究”国际学术研讨会

2017年11月10日-12日，语言工程与计算实验室云山青年学者刘伍颖副研究员赴南京参加第三届“面向翻译的术语研究”国际学术研讨会(NUTerm 2017)。

本次研讨会由南京大学、全国科学技术名词审定委员会和中国翻译研究院联合主办，邀请了来自术语学、翻译学、话语研究、信息技术等领域130多名中外学者围绕“术语·知识·话语”主题展开了跨学科、多视角的交流与对话，旨在进一步提升中国术语工作实践的国际化水平、促进中外跨语知识传播、加强中国学术话语权建设，同时推进中国术语学理论的发展。

刘伍颖副研究员应邀主持了“术语技术、教学与管理”分论坛，并在研讨会上宣讲了学术论文《基于语言大数据的韩汉术语库自动构建》，并展示了学术论文《国内译介研究术语演化分析》。

11. 实验室研究人员参加第十三届国际信息检索会议

实验室研究员任函博士参加了2017年12月5日-8日在日本举办的第十三届国际信息检索会议 NTCIR。该会议是世界三大著名信息检索评测会议之一，是相关领域的高水平会议。本届会议自动问答系统、短文本摘要、情绪原因分析等9个主题，共有71支队伍参加。课题组参加了短文本对话(STC)和情绪原因分析(ECA)两个评测任务，并在ECA任务中取得了单项评测第二的成绩。短文本对话和情绪原因分析都是自然语言处理中的核心任务，涉及篇章理解、语义推理和情感分析，因此极具挑战性。为此，课题组设计了各种统计模型，并通过反复实验确定了最优的模型组合及相关参数。

科研成果

2017 年度，语言工程与计算实验室专职研究员共发表三类以上论文 18 篇，新增立项国家级项目 1 项，省部级项目 3 项。新增发明专利 1 项，软件著作权 9 项，申请国家发明专利 2 项。

表 1. 科研项目

项目负责人	项目名称	项目来源	项目类别	项目级别
冯文贺	汉语篇章结构的特征—依存描写机制及资源建设研究	国家社科基金项目	国家级	一般项目
李键红	基于量化自相似性的图像超分辨率重建及评价算法研究	省科技计划项目	省部级	一般项目
杜家利	多维视角下的术语集约性研究	国家名词委项目	省部级	一般项目
杜家利	中国学习者英语花园幽径句的计算语言学研究	国家语委项目	省部级	一般项目

表 2. 论文及专利

序号	作者	论文名称	期刊名称	论文类别
1	刘伍颖	Supervised Ensemble Learning for Vietnamese Tokenization	International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems	一类 B
2	冯文贺	汉英篇章结构平行语料库的对齐标注评估	中文信息学报	二类 A
3	冯文贺	《世说新语》的篇章连接词	中文信息学报	二类 A
4	任函	基于语言现象的文本蕴涵识别	中文信息学报	二类 A
5	李霞	基于局部密度的无监督作文跑题检测方法	中文信息学报	二类 A
6	李霞	Automatic essay scoring based on Coh-Matrix feature selection for Chinese English Learners	SETE 2016	二类 A
7	李霞	Chinese Question Classification Based on Joint Semantic Features	第 6 届国际自然语言处理与中文计算会议/Springer	二类 A

8	杜家利	Controlling Interaction in Multilingual Conversation Revisited	Springer	二类 A
9	刘伍颖	POS-Tagging Enhanced Korean Text Summarization	ICIC 2017	二类 A
10	任函	Natural Logic Inference for Emotion Detection	Chinese Computational Linguistics and Natural language Processing Based on Naturally Annotated Big Data	二类 A
11	任函	Study on the English Corresponding Unit of Chinese Clause	Natural Language Understanding and Intelligent Applications	二类 A
12	任函	Sports News Generation from Live Webcast Scripts Based on Rules and Templates	Natural Language Understanding and Intelligent Applications	二类 A
13	李心广	An Approach to Evaluation Index and Model of Undergraduates' Spoken English Pronunciation	2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)	二类 B
14	刘伍颖	Unsupervised Ensemble Learning for Vietnamese Multisyllabic Word Extraction	IALP 2016	二类 B
15	刘伍颖	EFFICIENT KOREAN TEXT SUMMARIZATION BASED ON KEY PHRASE EXTRACTION	ICMLC 2017	二类 B
16	赵九茹	基于多属性效用理论的群体决策偏好整合研究	统计与决策	二类 B
17	赵九茹	The Text Mining Model Building of Open Questionnaire Based on LSA	IMCEC 2016	二类 B
18	刘伍颖	Vietnamese Multisyllabic-Word Extraction for Word Segmentation	International Journal of Asian Language Processing	三类
19	李心广	一种语音评分方法及系统	发明专利权	
20	李心广	英语口语发音评分方法和系统	申请国家发明专利并获国家专利局申请号	
21	李心广	基于移动互联网的汉英翻译教学评分系统	申请国家发明专利并获国家专利局申请号	

报送：广东外语外贸大学科研处

抄送：学校各科研单位

2017年12月31日印
