

# 语言工程与计算实验室

## 工作简报

(2014 年第 1 期 总第 1 期)

2014 年 9 月 30 日

---

### 新闻要目

1. 语言工程与计算实验室广外揭牌启动
2. 科研处召开专家研讨会把脉实验室研究方向规划
3. 孙茂松教授谈大数据下的语言信息处理
4. 周国栋教授围绕自然语言处理中的语义角色做精彩报告
5. 宗成庆教授做题为“统计学习方法在自然语言处理中的应用”学术报告
6. 实验室组织和申报广东高校国家级重点培育平台
7. 实验室老师喜获 2 项国家自然科学基金项目
8. 实验室老师积极参与和协助学校科研工作
9. 实验室老师指导本科学生获全国计算机设计大赛一等奖

## 要闻一：语言工程与计算实验室广外揭牌启动

2014年6月20日，语言工程与计算实验室在广东外语外贸大学南校区揭牌启动。省教育厅科研处副处长谢春艳，清华大学教授孙茂松，香港城市大学教授朱纯深，苏州大学教授周国栋，我校校长仲伟合、党委副书记陈建平、思科信息学院院长姜灵敏和副院长蒋益盛等出席启动仪式，仪式由我校副校长顾也力主持。



省教育厅科研处谢春艳副处长发言



仲伟合校长发言

为推动高校哲学社会科学与自然科学深度交叉融合，重点支持一批新兴交叉研究领域的哲学社会科学实验室，提升哲学社会科学的科学化水平，2013年我省批准了包括广东外语外贸大学语言工程与计算实验室在内的4个高校哲学社会科学重点实验室。

语言工程与计算实验室是我省高校中唯一与语言相结合的交叉学科重点实验室。未来语言工程与计算实验室将致力于语言学习、语言服务和信息理解三大领域的科学研究和实践，多学科交叉解决语言和社会科学问题，更好地服务广东经济社会建设，服务于国家语言文化战略发展。

启动仪式上，仲伟合用“三点”、“两地”、“一田”——“新兴交叉学科增长点”、“创新科研人才培育点”、“学校与社会重要联结点”，“重大项目孵化与产出地”，“重大项目

试验田”为实验室未来发展作出定位。仲伟合最后承诺，未来三年将投入 500 万经费用于实验室的硬件设备和科研，并预备申请国家级重点项目。启动仪式上，仲伟合还为实验室学术委员会委员颁发了聘书。



实验室启动仪式

实验室由信息学院牵头组织申报，并于 2013 年 12 月获广东省教育厅批准。2014 年 1 月 17 日，校长办公会决定成立语言工程与计算实验室，陈建平党委副书记兼任实验室主任。

实验室汇集了思科信息学院、英语语言文化学院、国际商务英语学院、东方语言文化学院、英语教育学院、中国语言文化学院、新闻学院、留学生教育学院、外国语言学及应用语言学研究 中心、词典学研究中心、翻译学研究中心的优势资源及海内外知名语言计算专家，依托广东外语外贸大学在语言学方面的领先优势，将多学科交叉解决语言和社会科学问题。实验室的主要研究方向包括：1) 语言计算基础理论与模型； 2) 语言知识库平台建设；3) 计算机辅助语言学习、测试、教学与翻译； 4) 社会计算、信息检索、文本挖掘及其在语言服务机制及信息处理平台中的研究。

实验室学术委员会成员分别为清华大学孙茂松教授、中科院自动化研究所宗成庆教授、苏州大学周国栋教授、哈尔滨工业大学刘挺教授、广东外语艺术职业学院曾用强教授、香港城市大学朱纯深教授、广东外语外贸大学陈建平教授、广东外语外贸大学刘建达教授、广东外语外贸大学姜灵敏教授，其中清华大学孙茂松教授担任学术委员会主席，广东外语外贸大学蒋盛益教授担任学术委员会秘书长。

## 要闻二：科研处召开专家研讨会把脉实验室研究方向规划

【主 持】陈建平

【时 间】2014 年 6 月 11 日 14: 10-15: 20

【地 点】北校区第一会议室

【出 席】陈建平、孙波、蒋盛益、温宾利、葛诗利、冯光武、姜灵敏、何国平、魏慧斌、李霞、柯晓华、李海琳

2014 年 6 月 11 日下午 14 时 10 分，在科研处的组织下，实验室负责人陈建平副书记、实验室组织申报单位信息学院相关老师蒋盛益、姜灵敏、李霞、柯晓华，以及来自学校英语语言文化学院、国际商务英语学院、东方语言文化学院、英语教育学院、中国语言文化学院、新闻学院、留学生教育学院、外国语言学及应用语言学研究、词典学研究中心、翻译学研究中心等合作单位的相关专家聚集在北校区第一会议室，讨论语言工程与计算实验室的研究方向规划问题，会议由陈建平副书记主持。

陈建平副书记首先发言，他指出，语言工程与计算实验室是我校第一个人文社科重点实验室，我校将会给予足够重视和支持。我校的语言学学科在国内不仅有较高地位，而且还很有特色。语言是由外语和中文构成，如何将我校的语言优势与信息科学研究优势结合起来开展多学科研究，是实验室未来的工作重点。陈副书记还指出，我们所成立的这个实验室不应仅仅看成是单一的实验室的概念，还应当成一个交叉学科来做，未来希望实验室能够成为我校新的科研增长点，能够产生出一批高显示度的成果出来。

科研处孙波副处长接着发言，他指出，2013 年 4 月份所发布的广东省哲学社科重点实验室项目，是广东省第一次发布的同类别的申报项目。我校采取 6+2 模式，即 6 个学院，2 个研究中心，共同申报该实验室，并获得成功。2013 年 12 月份，23 号文批准了该实验室的成立。该实验室将等同于校级人文社科重点基地，学校非常重视。从科研处的角度，希望实验室能够从研究方向，人员组成等方面取得突破，科研处将给与大力支持。

紧接着，信息学院蒋盛益副院长做了发言，他详细介绍了实验室的前期筹备工作，包括初步确定了实验室的研究方向、制定了实验室的规章制度、确定了实验室学术委员会的名单、确定了实验室的场地等。同时，他还大概介绍了实验室的设备方案、实验室的网站建设等。蒋盛益在发言中对在申报和筹备工作中来自各学院专家的意见、建议和支持表示了感谢，同时他还期望将来实验室的发展还需要大家的继续支持，希望各学院的专家和老师能够借助这样一个平台共同发展，共同进步。

接下来，来自各学院的专家老师，对实验室的研究方向和发展规划进行了深入的探讨和热烈的讨论。中文学院魏慧斌老师提出，语言工程与计算重点实验室对中文学院的老师来说是个

很好的机会。在他们的研究过程中，常常需要借助于某些工具软件，而有些软件非常昂贵，他们希望，能够借助实验室的平台和力量，对计算机辅助中文语言学习以及安卓平台上的汉语学习与应用等进行合作研究。来自英文学院的冯光武教授指出，应该在现有研究方向的基础上，继续凝练，最好以3-4个大方向为好，并有所突破，现有5个研究方向略多，并且实验室还应该在经费、场地以及人员配置上进行相对应的合理配置。来自新闻学院的何国平副院长则指出，实验室要办好，就应该是可持续发展的。实验室的高效率运转离不开专家导师以及在读博士和在读硕士的共同合作研究，因此，如何凝练方向，如何将交叉研究方向的硕士点、博士点以某种方式确定下来，如何将硕士生和博士生的招生融入到实验室的常规人才培养体系中来，是实验室可持续发展的一个非常重要的方面。会上，温宾利教授、葛诗利教授、姜灵敏教授等均做了发言。会议最后还确定了重点实验室的启动仪式时间，以及实验室定期举行学术前沿讲坛等运营模式。

### 要闻三：孙茂松教授谈大数据下的语言信息处理

2014年6月20日上午，实验室学术委员会主席清华大学孙茂松教授围绕大数据下的语言信息处理给广外师生做了一场精彩的报告。孙茂松就国内外的语言信息处理现状和趋势做了全面的介绍，通过大量的案例和数据深刻地分析了大数据背景下语言信息处理的必要性和重要性。面对国家对此的重大需求，他为语言工程与计算实验室指出了三个重要的研究领域，即深度学习、开放域信息抽取和知识图谱，由此思考了在模型创新和句子结构分析上的可能性，同时也指出核心问题是开放域句子结构分析。他点明了语言计算的本质是结构预测，即使在目前中文句法语义分析还较差的情况下，从大数据视角把现有技术、资源整合拼装起来，找到合适的高层次的交叉学科人才来完成这个经典又困难的任务是很有可能的，这充分要求科研人员们应该敢想敢做，在困难任务面前调动起全部神经。最后，孙茂松还介绍了借助互联网的中文研究与中文应用，充分展现了计算机与语言学交叉学科的魅力，对语言信息处理打开一个多彩而广阔的世界。



孙茂松教授谈大数据下的语言信息处理

附：孙茂松，清华大学计算机科学与技术系系主任，教授，博士生导师。研究方向为自然语言理解、中文信息处理和Web智能。作为项目负责人，主持973二级课题、863重大专项二级课题、国家自然科学基金重点项目、国家自然科学基金项目、863项目、国际合作项目等约20项，主持信息处理用分词国际标准2项。在国际刊物、国际会议、国内核心刊物上共发表论文130余篇，获得国家发明专利4项。多次担任相关领域国际会议和全国性学术会议大会主席或程序委员会主席。

## 要闻四：周国栋教授围绕自然语言处理中的语义角色做精彩报告

苏州大学周国栋教授于实验室启动当天为广外师生展开了主题为 Semantic Role Labeling for Natural Language Processing（自然语言处理之下的语义角色）的精彩报告。周教授借助“语义角色标注”和“中文语义分析中存在的困难”等例子，就“什么是自然语言”、“什么是自然语言处理”等做出详细的解释。作为一个音、形、义相结合的符号系统，自然语言用于人类的思维和交流，同时，自然语言也是一个变动的结构，由多部分组成。在中文语义标注问题上，存在着对语义、句法和篇章的分析并优化的困难，因此在自然语言处理过程，如何从细节入手，将文本信息转换为结构信息、将信息储存与交流平台开发为知识挖掘平台，于小处见大处，在研究领域也将得到越来越多的关注。



周国栋教授作学术报告

附：周国栋，苏州大学特聘教授（全职），博士生导师，目前主要从事自然语言理解基础研究和多语言跨文本信息抽取的研究。近5年来发表国际著名期刊和会议论文50余篇，包括 Bioinformatics(SCI 影响因子 6.7)、BMC Bioinformatics(SCI 影响因子 5.3)、Information Processing and Management(3 篇、SCI 影响因子 1.5)、Natural Language Engineering、ACL(8 篇)、EMNLP、COLING(4 篇)和 IJCNLP(4 篇)，其中 SCI 索引 20 余篇。担任过许多著名的国际杂志和会议的评审和委员会委员，包括 BMC Bioinformatics(SCI 杂志)，ACM Transactions on Asian Language Information Processing(ACM 杂志)。

## 要闻五：宗成庆教授做题为“统计学习方法在自然语言处理中的应用”学术报告

2014年6月19日下午14:15-16:15, 中科院自动化研究所宗成庆教授在南校区学术交流中心第一会议室展开了“统计学习方法在自然语言处理中的应用”学术报告。宗成庆教授首先提出自然语言处理所面临的挑战: 普遍存在的不确定性、未知语言现象的不可预测性、始终面临的数据不充分性、语言知识表达的复杂性等。然后介绍了统计学习方法的基本模型, 还介绍了统计学习方法在自然语言处理中的应用实例, 其中包括汉语自动分词、语块识别、语义角色标注和篇章分析等, 最后给出个人的一点思考和展望。宗教授指出, 网络大数据时代自然语言处理大有用武之地, 中文信息处理任重而道远, 基于统计学习的自然语言处理方法为我们提供了新的机遇的挑战, 面向汉语理解的计算模型和实现方法有待于更深入的研究和开发。



宗成教授作学术报告

附: 宗成庆, 教授, 1998年3月毕业于中国科学院计算技术研究所, 获博士学位, 1998年5月至2000年4月在中国科学院自动化研究所模式识别国家重点实验室从事博士后研究, 出站后留实验室工作至今, 现为模式识别国家重点实验室研究员, 博士生导师。

宗成庆博士于1999年和2001年两次在日本国际电气通信基础技术研究所(ATR)做客座研究员, 2004年在法国格勒诺布尔(Grenoble)信息与应用数学研究院(IMAG)机器翻译研究组(GETA)做短期高访。近几年来, 他主要从事自然语言处理理论与方法、口语自动翻译、人机对话技术等方面的研究, 目前担任亚洲自然语言处理联合会(AFNLP)执行理事、清华大学讲座教授、中国人工智能学会理事及自然语言处理专业委员会副主任、中国中文信息学会理事及机器翻译专业委员会副主任和若干国际、国内会议的程序委员会主席、程序委员会委员等多种学术职务。作为项目负责人承担国家自然科学基金项目、国家“863”项目、国家重大基础研究项目(973)子课题和国际合作研究项目等10余项, 近几年来, 在国内外重要学术刊物和会议上发表学术论文60余篇, 申请国家发明专利多项。

## 要闻六：实验室组织申报广东省高校国家级重点培育平台

经6月30日校长办公会讨论，根据申报要求，决定推荐语言工程与计算实验室为广东省高校国家级重点培育平台申报对象。

2014年7月2日上午9:00-12:00在北校三教422（中文学院会议室）举行了广东省高校国家级重点培育平台申报及建设研讨会，蒋盛益、王初明、杜金榜、袁传有、魏慧斌、葛诗利、刘梦莲、王凤兰、周咏梅、李霞、丘心颖、谢柏林、张新猛等参加了研讨会。会上针对实验室规划研究方向、实验室制度、研究成员及申报工作分工等进行了讨论。各位专家针对提出的初步研究方向提出了宝贵建议，最终初步确定了平台的研究方向及相关研究成员。

2014年7月12日上午9:00，学校针对“广东省教育厅关于组织广东高校国家级重点培育平台、重大培育项目与成果申报遴选工作”在北校区行政楼第三会议室进行专家指导答辩会，其中广东高校国家级重点培育平台-语言工程与计算实验室进行了答辩，蒋盛益教授在答辩会上介绍了申报材料，李霞、郝天永、丘心颖、张新猛等陪同参会，并根据专家提出的建议，对最终申报材料进行了修改、完善和补充，2014年7月14日正式提交申报材料。

实验室进一步厘清了主要发展方向为：1) 计算机辅助第二语言学习研究；2) 语言测试与评估技术；3) 面向安全战略的网络大数据语言应用研究。详细研究内容包括：计算机辅助第二语言学习模式研究、基于计算机网络技术的二语习得与语言新现象研究、基于语料库的外国学生汉语语块习得研究、满足模型校准等测试理论的测试题库建设、大规模考试中计算机辅助自动阅卷及自动评分研究、面向网络信息安全领域的信息抽取、法律信息抽取、社会情绪演化模型研究、基于网络语言理解的舆情研判与情感倾向性研究等。

## 要闻七：实验室老师喜获 2 项国家自然科学基金项目

---

2014 年度国家自然科学基金项目立项结果公布，我校喜获 8 项项目立项，其中面上项目 2 项，青年基金项目 6 项。

实验室老师李霞主持申报的青年科学基金项目“面向中国英语学习者的英文作文全自动评分及诊断反馈技术研究”、郝天永主持申报的青年科学基金项目“面向问答系统的多维语义模板研究”均成功立项，项目经费分别是 24 万元和 25 万元。

另外实验室老师王连喜所申报的教育部 2014 年人文社会科学研究青年项目“微博热点事件发现及其内容自动摘要研究”获得立项，项目经费 8 万元。

## 要闻八：实验室老师积极参与和协助学校科研工作

---

2014年7月，在阳爱民副校长的积极倡导下，实验室与外语研究与语言服务协同创新中心探讨合作研究外语教学与评估方面的软件开发，蒋盛益、郝天永、丘心颖、王初明、王凤兰、姜凌、林江豪、周剑峰等参与讨论，由郝天永负责研究现状的调研、系统的架构设计。

2014年8月，实验室协助外语研究与语言服务协同创新中心申报国家社科重大项目，由李霞和谢柏林分别撰写其中两个子课题。实验室与广东国际战略研究院探讨合作研究中国周边国家、海上丝绸之路相关国家的舆情和发掘战略信息。

## 要闻九：实验室老师指导本科学子获全国计算机设计大赛一等奖

在2014年（第7届）中国大学生计算机设计大赛决赛中，实验室教师蒋盛益、李霞、柯晓华指导的学生作品经过校、省级层层选拔入围全国总决赛。其中蒋盛益和柯晓华指导的作品于7月19-7月22日在沈阳东北大学参加了软件应用与开发类作品现场决赛，分别获得全国二等奖和全国三等奖。8月6日-8月9日，在杭州师范大学仓前校区举行了2014年（第2届）中国大学生软件服务外包类总决赛，李霞指导的作品“中英涉华网络新闻热点分析系统”得到评委的一致好评，获得全国一等奖。



一等奖获得者与指导老师合影

附：获奖作品

1. 中英涉华网络新闻热点分析系统. 郑景耀、姚昌宇、卢子晖、陈宏佳，指导老师：李霞，全国服务外包类一等奖.
2. 计算机专业英语个性化评测系统. 刘新月、陈映秀、姚星辉，指导老师：柯晓华，软件应用与开发类全国二等奖.
3. 基于微博的计算机学习资源自动分类系统. 周子程、黄卫坚、蔡茂丽，指导老师：蒋盛益，软件应用与开发类全国三等奖.
4. 班级小助手. 陈鹏州、邹元达、黄燕光，指导老师：蒋盛益，软件应用与开发类全国三等奖.



---

报送：广东外语外贸大学科研处

抄送：学校各科研单位

2014年9月30日印

---