

“小句复合体结构的理论与应用”学术研讨会

主办单位：广东外语外贸大学外语研究与语言服务协同创新中心

时间：2016年10月22日（星期六）8:30开始，为期一天。

地点：广东外语外贸大学（白云校区）六教 B108

“小句复合体结构的理论与应用”学术研讨会会议介绍

小句复合体(汉语中又称复句)是从小句过渡到语篇的关键节点。小句复合体的结构至少包括3个方面:小句之间词语共享关系,逻辑语义关系和指代关系。过去,语言学和计算语言学的研究和应用多侧重于小句。近年来语篇结构越来越受到重视,但这一层面的工作着重于逻辑语义关系和指代关系,少有关注小句之间词语共享关系的。其实,只有理清小句之间词语共享关系才能确定小句的构成,而只有确定了小句的构成才有可能进一步研究小句之间的逻辑语义关系和指代关系,才能服务于各种相关应用。

为了促进这一领域的工作,广东外语外贸大学外语研究与语言服务协同创新中心将举办“小句复合体结构的理论与应用”学术研讨会。

会议日程表

| 会议开幕式 | | | |
|-----------------------|---------------|---------------------------------|--|
| 时间 | 内容 | 负责人 | 备注 |
| 8: 15-8: 30 | 入场签到 | 简 阳 | 报名参加的老师和学生签到 |
| 8: 30-8: 40 | 会议开幕并 领导致辞 | 葛诗利 | 主持会议开幕并邀请领导致辞 |
| 8: 40-8: 50 | 合影留念 | 方 菲 | 摄影 |
| 会议主旨发言 | | | |
| 时间 | 发言人 | 主旨发言题目 | 备注 |
| 上午 (12: 00 结 束) | 葛诗利 | 英汉 NT 小句对齐语料库标注规范和进展 | 每位发言人的报告时间在 45 分钟左右, 发言结束后老师和同 学们可以积极提问, 发言中间穿插 休息时间, |
| | 丁建新 | 英语小句复合体投射系统之研究 | |
| | 荀恩东 | 汉语语块及其在句法分析中的应用 | |
| | 罗智勇 | 宾州英语树库到 NT 小句复合体结构自动转换的策略、方法和困难 | |
| 下午 (14: 00 开 始) | 尚 英 | 汉语语篇中背景句段的功能区分 | |
| | 蒋玉茹 | 基于语义的汉语新闻文本的零形回指消解研究 | |
| | 宋 柔 | 汉语语篇中广义话题结构与逻辑语义结构的关系 | |
| 会议自由发言 | | | |
| 时间 | 发言人 | 自由发言题目 | 备注 |
| 主旨发言之 后 | 孔繁丽 | 小句复合体中小句的关系; 汉语话题的句法语义表现 | 每位发言人有 15 分钟的 报告时间 |
| | 苏桡敏 | 美国总统辩论中的投射 | |
| 会议自由讨论 (30min-45min) | | | |

主旨报告 (1)

报告人：葛诗利（广东外语外贸大学）

时间：2016 年 10 月 22 日上午

地点：广东外语外贸大学（白云校区）六教 B108

题目：英汉 NT 小句对齐语料库标注规范和进展

摘要：基于微观话题和 NT 小句的概念，采用换行缩进的方法，英汉线性文本可表示成二维平面信息。我们以宾州树库原文的华尔街日报英语文本为原文语料，通过语料分析，探索英文语篇中的 naming 和 telling 的操作化定义。在语料分析和翻译过程中，构建了 NT 小句对齐语料库，并初步归纳总结了语料库标注规范。基于目前标注的英汉语料，我们归纳出由句子生成 NT 小句的 5 种模式。大量语料的标注实践说明，汉语和英语的篇章是按照 5 种模式生成的 NT 小句复合的结果，NT 小句的生成遵循保序性、不可穿越性，NT 小句具有成句性。NT 小句的这些性质可以从大脑认知机制的视角做出可信服的解释。

报告人简介：葛诗利，广东外语外贸大学教授，外国语言学及应用语言学、翻译学和商务英语研究硕士生导师，主要研究方向是计算语言学。目前主持研究包括国家社科面上项目、国家语委重点项目、教育部哲学社会科学一般项目、广东省社科项目等国家及省部级课题十余项，在自动作文评分、商务英语写作自动评价、机器翻译及翻译质量评价等方面取得成果。在语言学和计算机科学的重要刊物和会议上发表论文 30 余篇，多次获得广东外语外贸大学科研奖励。

主旨报告 (2)

报告人: 丁建新 (中山大学)

时间: 2016 年 10 月 22 日上午

地点: 广东外语外贸大学 (白云校区) 六教 B108

题目: 英语小句复合体投射系统之研究

摘要: 本文在 Halliday (1994) 研究的基础上, 对英语小句复合体投射系统进行重新构建。除确认并列、主从、嵌入三个子系统以及若干具有更大精密度的第二、三级子系统以外, 还对一些处在系统边缘的拟投射现象作了深入的讨论。同时, 文章还考察了“或然语法”(probablistic grammar) 中投射系统的运作情况, 将量化描写引入语法研究, 并用 Nesbitt & Plum (1987) 的“语义生成”(semo genesis) 假设加以解释。文章最后还调查了评论、报道、广告 3 种语篇体裁共 30 个语篇中投射型小句复合体的分布状态, 揭示了投射的语篇体裁变差, 并用语篇结构与语篇体裁之间的相互预测关系作出了初步的解释。

报告人简介: 丁建新, 中山大学外国语学院英语系教授、博士生导师。入选“教育部新世纪优秀人才支持计划”。兼任中山大学语言研究所所长、中山大学南方学院外国语言文学系主任。主要研究方向涉及批判语言学、体裁分析、功能语言学、文化研究。迄今在国内外重要学术刊物上发表论文近 40 篇。主持完成国家哲学社科项目 1 项, 参与国家重大哲学社科项目 2 项, 主持完成教育部、广东省项目 4 项。正主持编写全国英语专业“博雅系列”教材(总主编, 共 35 种)、“南开话语研究系列”(总主编之一, 10 余种)、Journal of Cambridge Studies 编委、《中国话语研究》编委、《话语研究学刊》编委。

主旨报告 (3)

报告人：荀恩东（北京语言大学）

时间：2016 年 10 月 22 日上午

地点：广东外语外贸大学（白云校区）六教 B108

题目：汉语语块及其在句法分析中的应用

摘要：语块是语言习用性的重要体现，在语料中高频出现，由连续或不连续的词语或语素整合而成，具有模块化、预制性、形式和意义一体化特征的实体性语言交际单位。本报告探讨汉语语块的几个问题：
1. 建立形式可区分的、面向计算的汉语语块分类体系；
2. 利用大数据，采用语言变换方法，提取各类语块实例；
3. 语块数据在汉语句法语义分析中的应用。

报告人简介：荀恩东，北京语言大学教授，北京语言大学大数据与语言教育研究所所长，毕业于哈尔滨工业大学，主要从事自然语言处理、语言教育技术研究和开发工作。

主旨报告 (4)

报告人：罗智勇（北京语言大学）

时间：2016 年 10 月 22 日上午

地点：广东外语外贸大学（白云校区）六教 B108

题目：宾州英语树库到 NT 小句复合体结构自动转换的策略、方法和困难

摘要：英语 NT 小句的分析，是 PTA (Parsing-Translating-Assembling) 英汉机器翻译模型中的首要步骤。标注具有一定规模和精度的 NT 小句复合体结构英语语料是进行机器自动分析和训练的基础。我们调查和分析了宾州树库句法结构标注体系和 NT 小句的 5 种模式之间的形式对应关系，构建从宾州英语标记树到 NT 小句复合体结构的自动转换系统。系统以宾州标记树中词形、节点标签、树形结构关系等作为转换特征；并采取转换规则和转换程序分离的策略，在人机交互过程中逐步提高转换的精度和一致性。我们也将分析 NT 小句复合体标注策略与宾州树库标注体系的区别，以及因此带来的转换困难。

报告人简介：罗智勇，北京语言大学信息科学学院副教授，硕士研究生导师，曾为美国北卡罗来纳大学统计与运筹系访问学者。作为主要研究人员，曾参加多项国家 863、国家自然科学基金项目；另主持教育部科学技术研究重点项目、北京市哲学社会科学规划研究基地项目和多项横向合作项目。参与和主持开发的现代汉语通用分词系统、计算机辅助汉语校对系统、全文检索系统、外媒报道分析系统等取得国内外领先或国际先进水平的成果，发表论文十余篇，并取得多项软件著作权，多次向国内外著名企业和研究单位转让技术研发成果，其中《现代汉语通用分词系统及其应用》2005 年获教育部提名国家科技进步二等奖（第二完成人）。

主旨报告 (5)

报告人：尚英（北京语言大学）

时间：2016年10月22日下午

地点：广东外语外贸大学（白云校区）六教 B108

题目：汉语语篇中背景句段的功能区分

摘要：广义话题结构理论是具有形式化体系、面向计算机应用的汉语语篇研究理论。背景句段在以往语言学的研究中或是句内的状语成分或是状语从句，而汉语中作为状语的背景句段同状语从句并无明确界限。本文在广义话题结构理论体系下，在一定规模语料库中，对背景句段进行了全覆盖的调查，对背景句段按照背景关系词及关系内容进行分类，放置在 NM 和 TL 的二元功能范畴中进行功能的区分。对于具有双重功能的背景句段，优先考虑其 TL 功能。对语篇中的标点句进行 NM-TL 分析，基本完成了语篇中实体关系分析，是逻辑关系分析的基础。此外，NM 与对应的 TL 组合成 NT 小句，模拟了人对语篇的理解程序，可以在机器翻译、文本摘要等应用方面发挥重要作用。

报告人简介：尚英，北京语言大学讲师，语言学及应用语言学博士，近年研究方向为汉语篇章语法，目前主持教育部社科青年项目“汉语篇章广义话题结构实证性研究”。

主旨报告 (6)

报告人：蒋玉茹（北京信息科技大学）

时间：2016年10月22日下午

地点：广东外语外贸大学（白云校区）六教B108

题目：基于语义的汉语新闻文本的零形回指消解研究

摘要：零形回指是汉语的一种典型现象，零形回指消解的研究对篇章分析、自动文摘、信息抽取、机器翻译等应用领域都将产生重要贡献。现有的零形回指消解研究大多利用先行词和回指词句法特征的一致性进行消解，一般需要依赖人工标注的句法树库。这是现有零形回指消解方法的局限。本项目将基于广义话题理论，在以往对话题句识别研究工作的基础上，提出两个基于语义的特征：一是候选先行词和回指词所在标点句的词汇语义组合可能性，二是候选先行词和回指词所在标点句构成的简单句的语义合理性，并将它们作为筛选候选先行词的两项重要指标。本研究将构建融合现有语义知识库和本体的词汇语义关系计算方法和基于大数据的基于统计的词汇语义相似度计算方法，并基于此构建词汇语义组合可能性评价模型；然后从大规模语料中抽取句子语义关系模式库，构建句子语义合理性评价模型；最后建立面向汉语新闻文本的零形回指消解研究系统，为机器翻译等与语篇理解相关的应用奠定基础。

报告人简介：蒋玉茹，北京信息科技大学计算机学院讲师。研究方向为自然语言处理、数据挖掘。目前已发表科技论文7篇，其中EI检索1篇，中文核心6篇；获得软件著作权2项。近5年参加国家自然科学基金3项，主持北京市教育委员会科研计划项目1项，其他项目3项。

主旨报告 (7)

报告人：宋柔（广东外语外贸大学）

时间：2016年10月22日下午

地点：广东外语外贸大学（白云校区）六教B108

题目：汉语语篇中广义话题结构与逻辑语义结构的关系

摘要：语篇的基本单位是小句，小句复合成小句复合体，小句复合体连成语篇。小句复合体的结构至少具有3个维度：话题共享、逻辑语义、成分同指。汉语话题共享结构即广义话题结构，其主体遵从流水模型。汉语小句复合体的逻辑语义结构，包括独立小句、序列关系和主从关系，主从关系中包括环境、辅助、原因、目的、条件、前提等多种二级关系。语料考察结果说明，广义话题结构内部小句间的逻辑语义关系综错复杂，但汉语广义话题结构作为一个整体与逻辑语义结构在绝大多数情况下是相容的。据此，可以将汉语语篇的小句复合体定义为以广义话题结构为主体的结构。如果两个相邻的广义话题结构具有紧密的主从关系，则它们合成一个小句复合体。除此情况之外，每一个广义话题结构是一个小句复合体。

报告人简介：宋柔，北京语言大学教授，广东外语外贸大学云山讲座教授，计算机应用专业和应用语言学专业博士生导师。近十几年的主要研究方向是汉语信息处理，先后主持十余项国家课题，在篇章分析、自动分词、机助校对、词语属性检索、字形计算、词类研究等方面取得成果。在计算机科学和语言学的重要刊物和会议上发表论文100余篇，研制并技术转让计算机软件多项，获专利2项，获北京市和教育部科技进步奖。目前主要研究方向是小句复合体的构成和计算。

广东外语外贸大学校园示意图及交通路线指引

广东外语外贸大学 校园示意图（北校区）

