

清华大学数字人文与 文学研究国际工作坊

清华大学中文系

芝加哥大学 Text Lab

6月10-11日

新斋 353

清华大学中文系

中国 北京

6月10日

8:00-8:10	刘石	会议致辞
8:10-8:30	全体合影	
8:30-9:30	霍伊特·朗	机器学习与文学模式辨认
9:30-9:50	戴梦菲	上海市图书馆民国文学期刊的电子数据化
9:50-10:00	茶歇	
10:00-12:30	苏真, 朱远骋 戴安德	以数据重新思考中国现代文学
12:30-13:30	午餐及休息	
13:30-13:50	戴安德	数字人文前史与早期清华大学文史研究
13:50-15:20	朱远骋	数字人文研究中的统计学方法
15:20-15:40	茶歇	
15:40-16:10	江铭虎	大数据环境下的语言认知与计算
16:10-17:00	霍伊特·朗	晴天日本文学数据库研究中的复杂关键词搜索

6月11日

8:00-9:00	霍伊特·朗	现代日本和中国小说中作为一种文体形式的重复
9:00-9:30	陈静	中国数字人文跨学科研究现状
9:30-10:00	但汉松	数字化文学批评的进路与反思
10:00-10:10	茶歇	
10:10-10:40	赵薇	社会网络分析与叙事性作品研究
10:40-11:00	严程	女性的友谊与诗：顾太清和她的秋红吟社诸姊妹
11:00-11:20	涂梦纯	《盗墓笔记》与《鬼吹灯》的计量风格学研究
11:20-12:00	戴安德, 姜文涛 陆晓芳	工作坊缘起, 数字人文与学术期刊合作, 及未来工作展望圆桌讨论
12:00-12:10	王中忱	会议总结及展望

引言

随着 20 世纪 90 年代以来信息技术的发展，计算的方法逐渐深入人文学科领域，并悄然改变着其研究方法。各种电子资料库、数据库等的建立和阅览检索方式的改变，使对于信息与技术、媒介之间关系的研究日渐重要，这正在改变传统意义上的以纸质媒体为主要对象、细读为主要方法的人文学科研究，也促使人们更加注意知识的形成及流传方式的变化。作为一种研究方法，“数字人文”针对文本现象进行量化分析，大致起源于 19 世纪的计量文献学、20 世纪 20 年代的俄罗斯形式主义，以及法国的年鉴学派。50 年代以来，这种研究方法的范围不断扩展。清华大学自 20 世纪 20 年代梁启超先生倡导“统计历史学”以来，就有人文社会科学量化分析的传统，此次我们邀请在数字人文研究方面初有成效的文学、语言学、统计学等学科的学人聚集，探讨新的技术和媒体环境下各个学科之间融合的可能性，希望能为人文社会科学的进一步发展有所贡献。

工作坊召集人

王中忱，刘颖，戴安德，姜文涛，赵薇

人员简介

霍伊特·朗

Hoyt Long

美国芝加哥大学东亚语言与文化系副教授，主要研究方向为现代日本文学，媒体史，传播学史，计算人文，地域文学，出版史，文学社会学，环境史等。著有 *On Uneven Ground: Miyazawa Kenji and the Making of Place in Modern Japan* (2012)，与苏真一起合作负责芝加哥文本 Text Lab。

苏真

Richard Jean So

美国芝加哥大学英文系副教授，主要研究方向为 20 世纪美国文学，美国亚裔文学，跨太平洋地区文学和文化，20 世纪中国文学，数字人文，媒体研究数等。著有 *Transpacific Community: America, China and the Rise and Fall of a Global Cultural Network* (2016)。

朱远骋

美国芝加哥大学统计学博士，宾夕法尼亚大学沃顿商学院统计系研究员，芝加哥大学 Text Lab 统计学专家。

戴安德

Anatoly Detwyler

美国哥伦比亚大学哲学博士，宾夕法尼亚州立大学人文与信息文化研究中心(The Center for Humanities and Information) 博士后研究员，哥伦比亚大学亚洲语言与文化系教员。研究中国现代文学和数字人文。

王中忱

清华大学中文系教授，博士生导师，系主任。研究比较文学，东亚近现代文学与文化史。

刘石

清华大学中文系教授、博士生导师，人文学院副院长。研究唐宋文学和古典文献学。

江铭虎

清华大学中文系教授，博士生导师。研究语言认知和计算。

邓盾

美国威斯康辛大学麦迪逊校区语言学系博士，美国布兰戴斯大学计算机科学系博士后，清华大学中文系助理教授，博士生导师。研究生成句法和面向自然语言处理的语言资源建设。

陆晓芳

《山东社会科学》杂志编辑

陈静	南京大学中文系博士，美国莱斯大学（Rice University）博士后，南京大学艺术研究院副教授。文化和媒介研究，新媒体，数字人文。
但汉松	南京大学英文系博士，南京大学英文系副教授，南京大学高研院驻院学者（2015-2017），主要研究兴趣为 20 世纪美国小说，后 9·11 文学，世界文学与数字人文。
赵薇	清华大学-康奈尔大学联合培养博士，首都师范大学文学院博士后，研究兴趣为比较文学与二十世纪中国文学，数字人文。
严程	清华大学中文系博士生，美国杜克大学联合培养博士生，研究方向为中国古代文论，清代女性文学文献。
涂梦纯	清华大学中文系研究生，研究语料库语言学。
姜文涛	美国纽约州立大学哲学博士，浙江大学外国文学研究所教员。英国“漫长的 18 世纪”印刷文化及情感研究，文艺理论和比较文学。
刘颖	清华大学中文系教授，博士生导师。研究语料库语言学和计算语言学。

发言摘要

机器学习与文学模式辨认	将数字人文的计算方法与文本细读及历史主义并置，呈现的具体研究对象是英语语言中俳句的语言及其形式。通过展示计算方法、细读及历史主义的研究方法是如何依赖于不同形式的“文本模式辨认，”我们可以看到这三种研究方法之间具有本质性的相似之处，这会改变我们通常所持有的数字人文与旧有的文学分析方法完全不同的观点。
上海市图书馆民国文学期刊的电子数据化	介绍上海市图书馆目前正在进行中的一项大规模的、针对民国时期文学期刊所采取的全文电子数据化项目。这项高质量的数据库计划近年内完成，这会改变学者们由文本出发进行文学史研究的形态。
以数据重新思考中国现代文学	运用计算的方法，探索如何使用上海市图书馆电子化项目最新的数据库，这个数据库目前包括了现代时期以来的 70 家重要的文学期刊。针对这样的数据库，我们会提供一些前沿的文本分析方法，同时，也认为大规模的量化分析方法既会引发我们在一个更为宽泛的意义和规模上重新思考中国现代文学史，也会让研究者发现新的研究方向。
数字人文前史与早期清华大学文史研究	从文本中提取数据，并以量化的方式分析文学和历史并非我们现在的时代才具有的研究方法。事实上，1920 年代，梁启超和他的学生就已经在清华大学提倡这种方法了。通过追溯这段“远读”（distant reading）的前史，我们可以认识到早期清华大学及民国时期中国的知识思想史为我们当前人文社会科学研究可以提供的价值。
数字人文研究中的统计学方法	介绍文本挖掘所使用的一些主要方法，同时探讨如下问题：统计学知识的重要性，数据识读素养（“data literacy”），数字人文研究中的跨学科合作。

<p>大数据环境下的语言认知与计算</p>	<p>语言的 ERP 脑电认知研究，基于深度学习与量子计算的语言处理，基于脑分子机制的语言认知。</p>
<p>晴天日本文学数据库研究中的复杂关键词搜索</p>	<p>展示晴天日本文学数据库高级关键词搜索界面。此界面开发于美国芝加哥大学，演示了数字人文如何改变了我们使用文本的方式。</p>
<p>现代日本和中国小说中作为一种文体形式的重复</p>	<p>使用前沿量化分析方法，探讨现代东亚文学中近代意义上的第一人称叙事角和心理叙事的出现，及其与语言重复（linguistic repetition）、言文一致运动影响的写作（vernacular writing）之间的关系。通过认识到语言上的重复是文学文体的一种因素，我们可以看到计算方法在进行比较性质的、跨语际的文学分析方面的可能性。</p>
<p>中国数字人文跨学科研究现状</p>	<p>数字人文，作为一个独立的、有边界的领域的指称，在中国出现是最近几年的事情。但以语料库、GIS（HGIS）和历史资料数据库为代表、具有数字人文性质的研究在中国却已经有了多年的发展。此发言对数字人文在中国进行历史性图绘，并对其发展境遇及特点进行梳理。</p>
<p>数字化文学批评的进路与反思</p>	<p>数字人文在图书馆学、媒体研究、认知心理学和数字化史学方面体现出了无可比拟的优势。回顾基于语料库和话语分析的文学文体学几十年来的发展，对我们进行数字化文学批评具有建设性意义。对文学研究方法独特性的强调，避免了对文学文本作为普通的语言质料加以数字化处理，会使得数字化文学批评免于沦为一般的语料库分析方法的附庸。</p>
<p>社会网络分析与叙事性作品研究</p>	<p>社会网络分析是数字人文中一种支柱性的理路与方法。此发言通过对李劫人超长篇历史小说中通讯（话语）网络的建构、计算和分析，将发现与阐释、大数据与小样本、文本挖掘与算法应用、外部研究与内部研究、文学社会空间与情节一体式分析真正结合起来，为文学研究者提供了从虚构作品入手进行数字人文研究的一种有益的思路。</p>
<p>女性的友谊与诗：顾太清和她的秋红吟社诸姊妹</p>	<p>从女性间友谊的建立、波折、聚散等时间节点着眼，通过顾太清（1799-1876）及其闺友的诗词作品中对于她们交往的记录，展现 19 世纪知识女性友谊的一种样貌，并通过数字化的展示方式，更直观清晰地呈现顾太清在不同生命阶段中交往、唱和的情况，以期为有清一代知识女性间友谊的深入考察提供新的视角。</p>
<p>《盗墓笔记》与《鬼吹灯》的计量风格学研究</p>	<p>从计量风格学的角度，对《鬼吹灯》《盗墓笔记》这两部同一题材的网络小说进行对比。出版十余年来，这两部同样有关盗墓、探险的作品一直被读者们进行比较。而与传统文学评论不同，计量风格学通过客观的数据统计和分析、而不是主观的阅读感受对文本的风格进行研究，力求得到真实全面、客观可靠的结论。</p>