

文章编号: 1003-0077 (2011) 00-0000-00

基于概念特征的汉语交互类言说动词语义分析及同义词群的建构*

肖珊¹, 郭婷婷²

(1. 中国地质大学(武汉) 国际教育学院, 湖北省 武汉市 430074;

2. 武汉大学 文学院, 湖北省 武汉市 430072)

摘要: 人工智能的最重要内容之一是自然语言处理, 建设词汇语义知识库是语言信息处理技术取得进步的重要保障。目前国内外基于同义词群(Synset)的词网研究存在结构不甚严谨, 语义颗粒度不够精细, 应用范围比较有限的问题。以“词群—词位”变体理论为基础建立多维特征概念语义网(Multi-D WordNet)是对这些问题进行解决的有效尝试, 通过分析与描写基本词位(概念本体)与变体之间、变体与变体之间的关系, 找出概念属性的差异, 从而构建颗粒度更为精细的同义词群。言说动词的次范畴类“交互类”动词可作为研究个案, 初步用来探讨这类典型词的语义结构及词群系统的建构规律。

关键词: 同义词群; “词群—词位变体”理论; “交互类”言说动词

中图分类号: TP391

文献标识码: A

Semantic Analysis of Interactive Speech Act Verbs on Conceptual Feature and its Synset Construction

XIAO Shan¹, GUO Tingting²

(1. International Education College, China University of Geosciences, Wuhan, Hubei 430074,

China; 2. College of Chinese Language and Literature, Wuhan University, Wuhan, Hubei
430072, China)

Abstract: Nature Language Processing (NLP) is one of the most important research parts of Artificial intelligence. And the Words Semantic Knowledge Base (WSKB) construction is an important guarantee to make big progress on NLP. Nowadays, the domestic and abroad researches of network of words which based on Synset have several problems: Less stringent structure, rough semantic particle size, Limited range of applications and so on. To build up a type of Multi-Dimension Word Net based on the detailed description of the features of concepts may solve these problems. This type of WSKB used the ‘Synset-Lexeme Anamorphosis’ Method to analysis the relationship and distinctive features between basis lexeme and its concept anamorphoses. And it introduced the interactive speech act verbs based on characteristic sense analysis as examples, preliminary discuss the description of the lexical meaning structure and its synset construction rules.

Keywords: Synset; Evaluated Speech Act Verbs; ‘Synset-Lexeme Anamorphosis’ Method

1 引言

21 世纪的到来是信息化社会全面发展的时代, 网络信息作为当今社会极其重要的资源之一已渗透到人类生活中的方方面面, 可以说人们获取信息的途径和手段也越来越依赖机器, 借助自动化手段处理海量语言信息的社会需求则日益凸显, 如何更好地、更快地推动语言信息处理技术的发展是计算机和语言文字学家共同关注的焦点。

* 收稿日期: 定稿日期:

基金项目: 国家社科基金重大项目《基于本体演化的语义网模型研究》(11&ZD189); 湖北省教育厅人文社科项目《基于词汇语义网的留学生科技汉语词汇创新性教学研究——以地质学词汇为例》(13g048); 湖北省社科基金《面向人机问答系统的汉语问答模式建构》([2009]129)

作者简介: 肖珊(1983—), 女, 博士, 讲师, 主要研究方向为中文信息处理、对外汉语教学; 郭婷婷(1976—), 女, 博士, 副教授, 主要研究方向为现代汉语语法、语言应用与信息处理。

语义网是人工智能的体现，它的提出是设想构造一种能够辨认、识别和自动处理人类语言的智能网络，就是为了解决目前计算机不能理解网页内容的语义和网上有用信息查准率较低的问题。对于我国而言，就是要利用计算机对各种汉语信息进行的一种自动化处理。“中文信息自动化处理每提高一步，给我国的科学技术、文化教育、经济建设、国家安全所带来的效益，将是无法用金钱的数额来计算的。”^[6]

迄今为止，虽然中文信息处理取得了不俗的成绩，但是要想机器理解自然语言传递的信息，一个重要的前提就是要用最经济的模式告诉机器足够多的语言知识，语言资源建设尤其是词汇语义知识库的建设应运而生。综观现有的国内外的语义知识库比如 Word Net, Frame Net, CCD, HowNet, HNC 等^[4]，它们虽然从不同的角度对概念之间的语义关系进行了多方面的描写，但是对于整个概念系统内部的独立的个体即概念的变体的区别性语义特征挖掘力度还不够，语义颗粒度不够精细，因而词网的结构和应用覆盖面还有待进一步完善和拓宽。同时，词汇本身的处理技术的提高也会为句处理带来突破性的意义，为计算机的自动化“习得”起到巨大的帮助作用。

2 言说动词及其同义词群的建构理念

2.1 关于言说动词

言语活动是人们传递和交换信息的主要手段之一，映射到词汇系统上就是以“说”为代表的一系列言说动词。

早年对言说动词的研究是将其贯彻于其他动词研究的过程之中，比如在研究动词的价、词义辨析等时涉及，而那时有的学者研究的方式和角度很有限：或仅举动词实例来说明什么是言说动词，或只对言说动词的上位词如“说”等下定义；即便是从本世纪初开始对言说动词进行专项研究的钟守满^[9]、刘大为等学者，也并没有对言说动词有一个统一而全面的定义。我们认为言说动词的核心语义成分是围绕“言说”动作行为而联系的，基本的表现形式为(图1)：

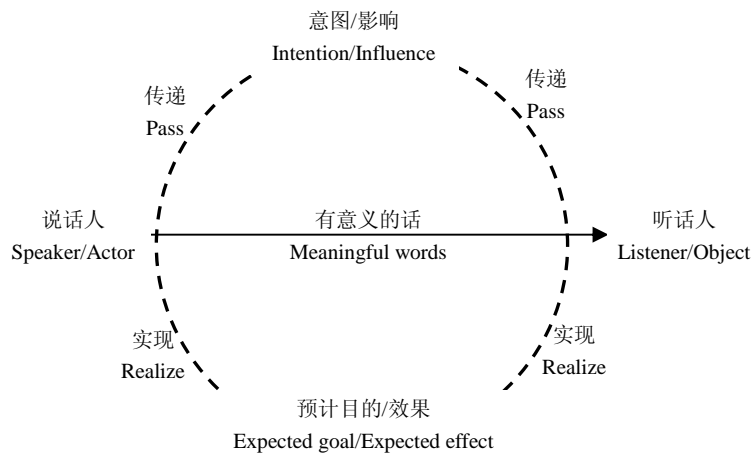


图1 言说核心语义成分表现示意图

因此可以定义言说动词为：说话主体用有效的语言符号表达某意图（显性或隐性）或向目标对象（人或事物）传递某信息，从而达到某目的或某效果的一类动词。

作为人类言语系统研究的重要组成部分，典型的言说动词的语义网络系统研究有助于从某个侧面或角度来探讨自然语言是如何生成和被理解的，帮助语言学家更准确地进行语义结构的形式化描写，也可以帮助计算机学者对动词系统乃至整个自然语言系统的计算机模拟起到启发和推导作用。

2.2 “词群-词位变体”理论的提出

从 2004 年起,以萧国政教授为首的武汉大学语言与信息研究中心师生们卧薪尝胆,开始了“基于概念特征属性的汉语词网建构”的艰苦工程及其指向语言处理的理论探索,“词群-词位变体”理论^[7]就是在实践研究中探索出的主要理论成果之一。理论的提出是基于人们认识事物的基本认知规律之上的。

人们认知事物时一般都会存在认知原型,认知的原型事物反映到词汇系统中就构成了原型基元词,这些基元词所在的位置就是基本词位(概念本体),其他的词汇都是以这些基元词为依托,通过对概念在不同维度上属性取值的变化而形成的变体,基本词位(概念本体)及其变体位共同构成词群。在词群中,词的概念义是词群的纲,它是把所有属于同一概念范畴的词统合在一起的根本依据;而词的同一概念义在词汇系统中又往往表现出不同的词语形式,这些不同形式的词语因具有词群的群义特征而相互联系,同时又因具有独特的个性义而相互区别。因此“词群-词位变体”理论的核心思想就是:一种语言的词义系统是以基本词位(概念本体)及其独特义的变体构成的同义词群。

通过对基本词位(概念本体)及其变体构成进行义素分析能显示一义跟他义的语法语义对立及联系,进而可以描述词语之间、词群之间纷繁复杂的语义、语法关系。这种思想方法在很大程度上能满足自然语言理解中对语义形式描写的细致要求,而归纳出细致准确的语言规则、组建大规模的丰富的语义知识库能帮助机器像人脑一样更科学更快地处理自然语言,实现人机合作人机对话。

2.3 研究对象的分类与个案确定

“言语性”要素是言说动词区别于其他各类动词的根本,那么根据语义结构中所凸显出“言语性”的不同方面,可将言说动词分为四类^[2],如下所示:

- 问答类:强调行为原因或条件。例如:询问、征询。
- 交互类:强调施动者数量或形态。例如:交谈、采访。
- 客事类:强调言语行为关系。按照强调的关系不同,又分为强调传递意见看法的“评价类”,例如:点评、评论;强调传递某情感的“情感类”,例如:道喜、唾骂;强调传递某信息的“告知类”,例如:通知、传达。
- 祈使类:强调为达到目的行为,例如:请求、劝说。

限于篇幅,本文选取以言说动词为对象的义合网络的构建的重要部分之一——“交互类”言说动词做个案分析。所有语料均源于《现代汉语词典》^[8],词语释义同时参考《现代汉语动词大词典》^[5]。

3 “交互类”言说动词

3.1 语义特征及再分类

“交互”类言说动词是一类反映说话者(主事)与听话者(客事)之间通过言语内容的互动达到交流信息或交换意见、看法的言语行为动词。它最显著的语义特征就是交互的“双向或多向性”,按照“交互”所指的不同,又可分为五类:(1)以交流信息、意思为目的的“交谈类”;(2)以交换意见看法为目的的“讨论类”;(3)以达成共识或解决问题为目的的“商量类”;(4)以侧重批判和揭露对方矛盾为主要目的的“辩论类”;(5)以侧重维护自我为主要目的的“争吵类”。用表(表1)可表示为:

表1 “交互”类言说动词分类表

语义分类项目	词语举例
交谈类	谈 ₁ ^① 、谈话、交谈、对话、倾吐、闲谈...
讨论类	讨论、议论 ₁ 、探讨、谈论...
商量类	商量、面商、磋商、相商、婉商、会商...
辩论类	辩论、争辩 ₁ 、理论 ₂ 、论难、论战、论争...
争吵类	争辩 ₂ 、争吵、吵闹 ₁ 、吵架、吵嘴、口角...

本文只探讨“辩论类”言说动词词群的“概念本体-变体”类型，期以一斑窥全豹。

3.2 “辩论类”言说动词的“概念本体-变体”语义类型

同义词群是由基本词位（概念本体）及其变体构成的。“概念本体-变体”的语义关系类型按照逻辑联系，又可分为“三种关系、四个方阵”，即“同位关系、上下位关系、邻位关系”及关系中基本词位及其变体所处的“同位(PS)”“上位(PU)”“下位(PD)”“邻位(PB)”四个位置。如图2所示：

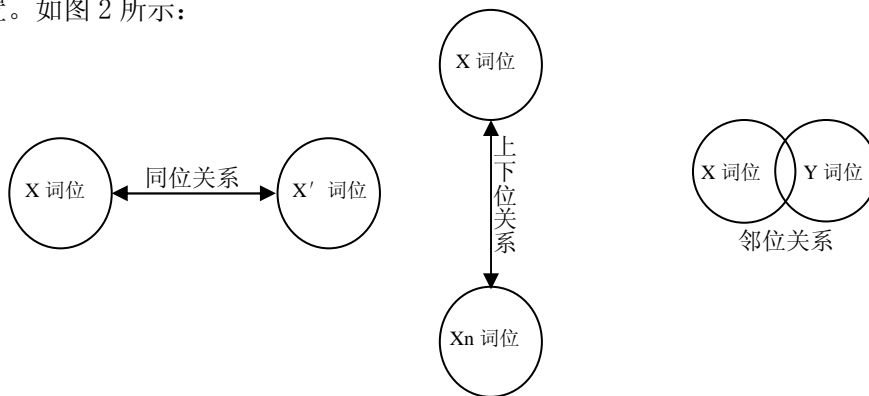


图2 “概念本体-变体”的语义关系类型图

以“辩论类”言说动词同义词群来说，它与整个“交互类”言说动词词群是上下位关系，而“辩论类”言说动词又因语义侧重点的不同可细分为不同的同义词群，每一同义词群内部又有其基本词位(概念本体)和变体，它们都处于或属于这“三种关系、四个方阵”之中。

1) 同位关系

同位关系是指词的概念本体(X)与其语义变体(X'PS)只有词语形式上的不同，语义内涵完全相同，在不同句法环境下可以相互替换，概念本体与变体之间的位置不分差别。

例如“辩论”“论辩”和“辩₂”都是“主事”和“客事”之间就同样的问题或话题，持续地用一定的、相互对立的理由来说明自己对事物或问题的见解，以揭露对方的矛盾的动作行为。它们的语义解释几乎等义，在不同句法环境下又可互换，唯一的区别就是由于“辩₂”的单音节局限而导致在句法语用上受到一定的限制，不能单独使用，必须添加动补成分，如：

①大家都为梅森论辩/辩论时所运用的那种高明手法而向他纷纷祝贺。

②两个人在那里为了一点意见不统一而辩论/辩个不停，最后竟大打出手，实在不可思议。

因此，三者的句法语义结构式同为：[动作/行为+揭露矛盾+交换看法/意见+持续+用对立理由+客事←→问题/话题←→主事]^②，三者互为同位变体，也可是说这三者组成的集合

^①这里的下标“1、2...”代表这个词在词典中的义项义。下文中若有类此情况，解释相同。

^②这里采用的是萧国政教授985二期拓展项目“基于概念特征的汉语语义网的理论探索与技术”(985yk006)课题组使用的句法控制下的义素分析法，结构式的理解顺序是从右到左进行解读。文中所有句法语义结构式的解读同理。

是同位变体词群。如图 3 所示：

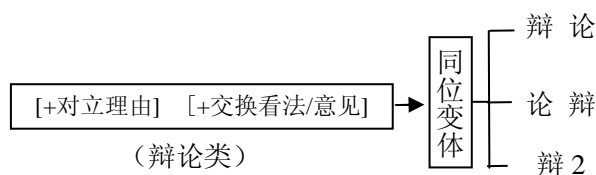


图 3 “辩论”类同位变体关系图

2) 上下位关系

上下位关系是指词的概念本体 (XPU) 与其语义变体 (X'PD) 是处于上下位置的概念, 概念本体与变体之间不仅词语形式不同, 而且变体的语义内涵蕴含于概念本体中, 是在概念本体内涵义基础上通过添加或改变不同的区别性语义特征而实现的。

同样是表示主事、客事之间就某一问题或话题交换看法或意见, 但因态度的不同, 所形成的下位同义词群群义所指就不尽相同。例如同义词群 {争持、争、争执、争竞} 侧重凸显“争”即【+相持不让】区别性语义特征而成为“辩论”的下位变体, 句法语义结构式为: [动作/行为+揭露矛盾+交换看法/意见+态度相持不让+持续+用对立理由+客事 ↔ 问题/话题 ↔ 主事], 而同义词群 {嚼舌²、抬杠¹} 则主要侧重凸显说话者的【+无谓/毫无价值】的主观态度, 句法语义结构式为: [动作/行为+揭露矛盾+交换看法/意见+持续+态度无谓+用对立理由+客事 ↔ 问题/话题 ↔ 主事+口语]; 再者, 由于主事、客事言说行为涉及的问题或话题范围的不同, 所形成的下位同义词群所指也会不同。例如同义词群 {论战、论争、争鸣} 所涉及的范围就缩小到【+政治/学术问题】, 句法语义结构式为: [动作/行为+揭露矛盾+交换看法/意见+持续+用对立理由+客事 ↔ 政治/学术问题/话题 ↔ 主事]; 另外, 强调主事、客事在交换看法或意见时所使用的的方式时, 所形成的下位同义词群也不一样, 如同义词群 {论难、辩难、舌战} 都是强调主事与客事之间用【+攻击方式】去表达自己的观点, 揭露对方的矛盾, 句法语义结构式为: [动作/行为+揭露矛盾+攻击看法/意见+攻击方式+持续+用对立理由+客事 ↔ 问题/话题 ↔ 主事]。如图 4 所示:

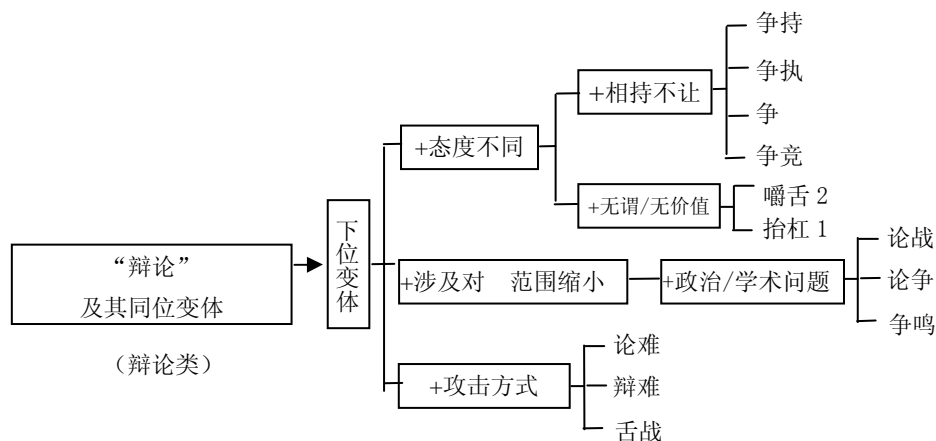


图 4 “辩论”类下位变体关系图

3) 邻位关系

上下位关系是指词的概念本体 (XPU) 与其语义变体 (X'PB) 是处于相邻位置的概念, 概念本体与变体的语义成分有一定交叉, 但变体同时又包含有其他的基本语义要素, 这些语义要素是变体本身不可缺少的。这些交叉有可能是同一类词群内部次词群之间言说动词义的交叉, 我们称之为“言语内交叉”, 也可能是不同类词群之间言说动词义的交叉, 我们称之为“言语外交叉”。

例如同义词群 {争议、争论、理论²、争辩、说嘴²}, 它们都包含【+用对立理由】【+交

换看法/意见】这样的语义，这是所有“辩论类”言说动词同义词群共有的核心语义，是它们属于“辩论类”言说动词同义词群的基本条件，但是我们同时还发现它们有属于其他同义词群的一些基本语义要素，比如“争议、争论”除了含有“辩论”义之外，还含有“争执”义即属于前述同义词群{争持、争、争执、争竞}中的区别性语义特征【+相持不让】，同时还包含“讨论”义，这是“讨论类”言说动词的基本语义要素之一，并且这三义并不存在侧重凸显，三者都是不可或缺、互相糅合的，加之“讨论类”也是交互类言说动词的一种，因此我们说这就属于“言语内交叉”；再如“争辩”“说嘴”也不单单是一种“辩论”，而是通过“辩论”的方式来强烈地表达自己观点的正确，是为保护自己或他人而进行的“辩护”，含有及强调【+解释/说明】的含义，这是“解说类”言说动词的基本语义要素之一，因此我们说这属于“言语外交叉”。

与此同时，我们也知道，语言词汇系统不会按照我们的想象而有序排列，也就是说总有一些词语在规则之外，比如在“辩论类”言说动词词群中还有一些词虽然是下位变体，但是它们是同时凸显两个或多个义征的，比如“辩难”不仅是凸显强调主事与客事之间用【+攻击方式】去表达自己的观点，更是强调用一种【+书面的】【+提难解问题的形式】去质疑对方的观点和看法，同时在【+强度/程度属性】上也更凸显【+激烈地】，因此它的句法语义结构式更为精细，为：[动作/行为+揭露矛盾+攻击看法/意见+激烈地+书面提问的攻击方式+持续+用对立理由+客事↔问题/话题↔主事]

对应到计算机对词汇语义的处理上，要激活某同义词词群中所有成员，可以找出原型基位词作为整个 Synset 的节点 (ontology)，其他的非原型基位词为概念的变体，在基位词属性基础上改变一个或几个属性特征即可。对于整个词汇语义网络建构来说，首先需要语言学家确定有限集合的原型基本词位及限定附属语义特征，而后再根据词汇生成的规律来告知计算机如何分解、比对及自动增减语义特征而生成不同概念变体，从而实现计算机处理词语技术速度和效率上的提升。

比如计算机想要识别基本词位“辩论”和它的一部分下位变体如“争执”“嚼舌”“论战”“论难”，就可以在找到及确认“辩论”这个节点之后，通过激活不同指向上的区别性语义特征，找出它们之间的联系。如图 5 所示：

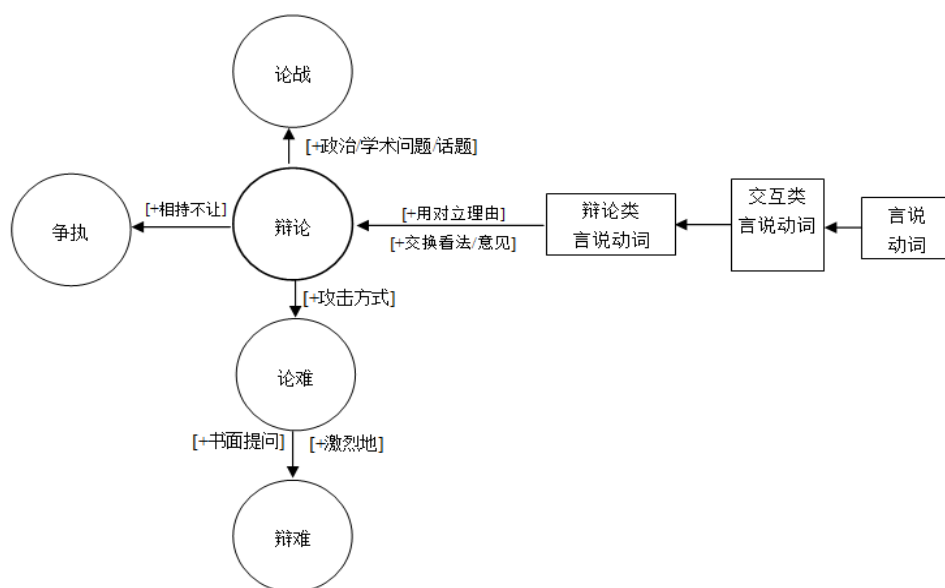


图 5 “辩论”及其部分变体的计算机识别示意图

4 结语

本文是以“词群—词位”理论为核心来对汉语语义网中交互类言说动词部分的同义词群的建构进行的一次探索。通过分析,我们可以看到以词群中的原型基本词位为核心,以改变各种附属语义特征的方式,能扩展出相应的同位、邻位和下位变体,语义之间的相关性是联系词与词之间的纽带。这样方法能弥补目前词网结构不严谨、语义颗粒粗糙等问题,能更好地帮助计算机对自然语言处理特别是语义识别上的实际需求。

虽然本文只是选择动词词群对汉语词汇语义网络模式的建构进行了尝试,但是对其他词类如名词、形容词甚至是虚词词群的语义网络模式建构有很好的启发价值,而整个汉语词汇语义网络就是由各类词群构成的系统性的网络整体。

参考文献

- [1] 常颖. 汉、俄语言语行为动词语义对比研究[D]. 吉林: 黑龙江大学, 2008.
- [2] 蔡俊杰. 现代汉语言说类动词考察[D]. 上海: 上海师范大学, 2008.
- [3] 胡焯. 基于多维特征属性描写的现代汉语概念语义网的建构研究[D]. 武汉: 武汉大学, 2007.
- [4] 黄曾阳. HNC(概念层次网络)理论——计算机理解语言研究的新思路[M]. 北京: 清华大学出版社, 1998:1-9.
- [5] 林杏光, 王玲玲, 孙德金等编著. 现代汉语动词大词典[Z]. 北京: 北京语言学院出版社, 1994.
- [6] 许嘉璐. 现状和设想——试论中文信息处理与现代汉语研究[J]. 中国语文, 2001(6): 2.
- [7] 萧国政. 动词“打”本义的结构描写及其同义词群建构——一种人机共享的“词群—词位变体”研究初探[C] // 萧国政, 姬东鸿, 孙茂松主编. 中文计算技术与语言问题研究——第七届中文信息处理国际会议论文集. 北京: 电子工业出版社, 2007: 4-7.
- [8] 中国社会科学院语言研究所词典编辑室编. 现代汉语词典(第5版)[Z]. 北京: 商务印书馆, 2005.
- [9] 钟守满. “互向”类言语行为动词语义认知解释[J]. 杭州师范学院学报(社会科学版), 2005(3): 109-112.

作者 1 联系方式:

姓名 肖珊

地址 湖北省武汉市洪山区鲁磨路 388 号中国地质大学（武汉）国际合作处

邮编 430074

电话（最好手机） 18971639105

电子邮箱 wdxshan@gmail.com / 84781725@qq.com

作者 2 联系方式:

姓名 郭婷婷

地址 湖北省武汉市武昌区珞珈山武汉大学文学院

邮编 430072

电话（最好手机） 18971639102

电子邮箱 wdguott@gmail.com