



浙江省哲学社会科学规划
后期资助课题成果文库

汉语零形回指认知的 多维研究

Hanyu Lingxing Huizhi Renzhi De
Duowei Yanjiu

王倩 著

中国社会科学出版社



浙江省哲学社会科学规划
后期资助课题成果文库

汉语零形回指认知的 多维研究

Hanyu Lingxing Huizhi Renzhi De
Duowei Yanjiu

王倩 著

中国社会科学出版社

图书在版编目(CIP)数据

汉语零形回指认知的多维研究 / 王倩著. —北京: 中国社会科学出版社,
2017. 5

(浙江省哲学社会科学规划后期资助课题成果文库)

ISBN 978 - 7 - 5203 - 0531 - 0

I. ①汉… II. ①王… III. ①现代汉语 - 词法 - 研究 IV. ①H146.1

中国版本图书馆 CIP 数据核字(2017)第 134896 号

出 版 人 赵剑英
责任编辑 宫京蕾
特约编辑 李晓丽
责任校对 周 昊
责任印制 李寡寡

出 版 **中国社会科学出版社**
社 址 北京鼓楼西大街甲 158 号
邮 编 100720
网 址 <http://www.csspw.cn>
发 行 部 010 - 84083685
门 市 部 010 - 84029450
经 销 新华书店及其他书店

印刷装订 北京市兴怀印刷厂
版 次 2017 年 5 月第 1 版
印 次 2017 年 5 月第 1 次印刷

开 本 710 × 1000 1/16
印 张 12.75
插 页 2
字 数 213 千字
定 价 56.00 元

凡购买中国社会科学出版社图书, 如有质量问题请与本社营销中心联系调换

电话: 010 - 84083683

版权所有 侵权必究

目 录

第一章 绪论	(1)
第一节 选题意义	(1)
第二节 相关研究综述	(4)
第三节 研究内容与研究方法	(11)
第四节 本章小结	(13)
第二章 代词脱落的跨语言视角及汉语零形回指的分类与分布	(14)
第一节 代词脱落的跨语言视角	(14)
第二节 汉语零形回指的分类与分布	(25)
一 汉语零形回指的分类	(25)
二 汉语零形回指分布的语料库统计	(27)
第三节 本章小结	(31)
第三章 回指研究的相关理论	(33)
第一节 可及性理论	(33)
一 可及性理论的内容	(33)
二 可及性理论在汉语零形回指研究中的应用	(35)
第二节 向心理论	(37)
一 向心理论的内容	(37)
二 向心理论在汉语零形回指研究中的应用	(39)
第三节 认知参照点理论	(43)
一 认知参照点理论的内容	(43)
二 认知参照点理论在回指研究中的应用	(44)
第四节 语篇表征理论和文本更新理论	(45)
一 DRT 和 FCT 的内容	(46)
二 DRT 在指代研究中的应用	(48)

第五节 本章小结	(49)
第四章 汉语零形回指认知的先行语凸显因素	(51)
第一节 语法显著	(53)
一 语法显著序列	(53)
二 语法显著序列的认知心理学解释	(57)
三 语法显著与回指策略	(61)
四 “主语”和“主题”之争	(63)
第二节 语义显著	(66)
一 论元的语义角色	(67)
二 名词生命度与语义显著	(72)
三 动词施事性与语义显著	(76)
四 构式与语义显著	(80)
第三节 信息显著	(87)
一 信息结构理论	(88)
二 无定名词与信息显著	(89)
三 动词语义与信息显著	(95)
第四节 本章小结	(103)
第五章 汉语零形回指认知的结构因素	(106)
第一节 汉语零形回指认知的距离因素	(106)
一 工作记忆与认知负荷理论	(107)
二 线性距离与认知距离	(108)
三 距离因素与零形回指照应模式	(109)
第二节 复句语义关系因素	(120)
一 修辞结构理论与汉语复句类型	(120)
二 复句语义关系类型与零形回指先行语语法角色	(122)
三 复句语义关系的形式标记	(124)
第三节 本章小结	(129)
第六章 实验	(131)
第一节 实验1: 先行语位置策略实验	(132)
一 实验方法	(132)
二 实验结果	(134)
第二节 实验2: 主语显著的共性实验	(135)

一 实验方法	(135)
二 实验结果	(137)
第三节 实验3: 名词生命度实验	(138)
一 实验方法	(138)
二 实验结果	(139)
第四节 实验4: 动词语义和复句语义关系实验	(140)
一 实验方法	(140)
二 实验结果	(142)
第五节 名词有定性实验	(144)
一 实验方法	(144)
二 实验结果	(146)
第六节 数量信息实验	(147)
一 实验方法	(147)
二 实验结果	(149)
第七节 距离因素实验	(150)
一 实验方法	(150)
二 实验结果	(154)
第八节 讨论	(155)
第九节 本章小结	(159)
第七章 汉语零形回指的生成与解析机制	(160)
第一节 汉语零形回指的生成机制	(161)
第二节 汉语零形回指的解析机制	(163)
第三节 本章小结	(165)
第八章 结语	(166)
第一节 主要特点	(166)
第二节 不足之处	(167)
第三节 研究展望	(168)
参考文献	(170)
附录	(184)

第一章

绪 论

第一节 选题意义

在语篇中，我们经常会看到这种现象：某个成分（包括人物、事物、事件、概念、现象等）在语篇中如多次被提及，作者有可能重复使用这个成分，也可能不再重复使用这个成分，而是使用另一个成分来指称这个成分，这种现象被称为指代（anaphora），在很多研究中也称为回指。关于指代，语言学家们给出了很多不同的定义，其中最为经典的是韩礼德和哈桑（Halliday & Hasan, 1976）提出的将指代看作一种实现篇章衔接的语言现象。对于指代研究的价值和意义，黄衍（Huang, Y., 2000）指出，对于指代的研究有助于我们回答乔姆斯基所提出的语言的本质的问题——“柏拉图问题”的一个特例，即语言习得的逻辑问题。由于当代语言学把指代看作一种心理表征（representation），因此，对指代现象的研究是帮助人们加深理解人类心智本质的最有效的途径之一。长期以来，人们从哲学、逻辑学、语言学、计算科学、人工智能等众多领域对指代现象展开了多角度、全方位的研究。在语言学层面上，学者们从句法学、语义学、语用学、语言认知、篇章分析、文体学、语言类型学和语言普遍现象研究、语言习得和语言教学等不同角度，对语言中的指代现象进行了广泛研究。其中20世纪80年代是回指研究的爆炸期，在此之后，回指研究便发展成为理论语言学关于句法、语义和语用界面上的一个中心研究课题。

随着语言学和心理语言学对语篇理解或处理的日益重视，指代消解也成为该领域的一个研究热点。所谓“指代消解”（anaphora resolution），就是为回指语（anaphor）确定先行语（antecedent），也就是其指代对象的过程。王军（2005）指出，回指研究的核心问题是回指释义，即回指语

在各种因素的作用下搜寻到其先行语的过程。在这一过程中,各种各样的语言和非语言因素都可能牵涉其中,而回指研究的任务就是要把这些因素剥离出来,考察它们是如何对回指释义的过程产生影响的。

指代消解是自然语言处理的关键和热点问题之一,是信息抽取不可或缺的组成部分,在自然语言的篇章理解中具有举足轻重的意义。为了更准确且没有遗漏地从文本中抽取相关信息,必须要对文章中的指代现象进行消解。指代消解不仅在信息抽取中具有重要的价值,而且在自然语言接口、机器翻译、文本摘要和问答系统等应用中也起着关键的作用。因此,指代消解研究可以极大地推动自然语言处理研究的发展,促进自然语言处理研究与应用의 紧密结合,具有非常高的应用价值。

国际上关于指代的既往研究主要集中在以下三个领域:人工智能、经典语言学、认知心理学。许多早期的方法侧重于从理论上进行探索,运用大量手工构建的语言甚至领域知识进行指代消解。近年来,由于自然语言自动处理技术的发展以及各类应用对指代消解技术的需求越来越迫切,人们转向了基于弱语言知识的方法,将研究重点放在了实用的自动指代消解技术的研究开发上,如人工智能领域的机器学习的方法等,并取得了一定的进展。然而,受制于弱语言知识,自动指代消解技术近年来在性能的继续提高上遇到了“瓶颈”。无论是回指的生成,即说话人在先行语候选集合中挑选出最合适的先行语后,在下文中用某种回指形式和该先行语进行照应,还是回指的解析,即听话人在接收到回指语时即时搜索其相应的先行语,从而确定回指语的所指对象,在这一过程中,语言的使用者都需要激活头脑中相关的句法信息、语义信息和百科知识等,根据各种限定性条件确定回指形式或回指语对应的先行语。而这些深层语言知识,特别是结构化句法信息和语义信息等知识都是机器所无法识别和理解的。因此,要在指代消解技术上突破“瓶颈”,还需要语言学领域的研究者从深层语言知识的层面为指代加工的认知机制提供更多的理论支持和规律破解。

英语中最常见的回指形式主要是代词回指,因此在西方语言学家对回指的研究中,最受关注的是代词回指。和英语不同的是,汉语中的零形回指的现象非常普遍。所谓零形回指,按照霍珀和汤普森(Hopper & Thompson, 1980)的解释,指的是失去的论元可以没有变化地在语法上补出来。陈平(Chen, P., 1987)是这样界定零形回指的:如果从意思上讲句子中有一个与上文中出现的某个事物指称相同的所指对象,但从语法

格局上看该所指对象没有实在的词语表现形式,我们便认定此处用了零形回指。这两个对零形回指的定义从本质上来说是一样的,只不过霍珀和汤普森是从动词的论元的角度来谈的,而陈平是着重从上下文情景的角度来思考的。把这两个定义结合起来,所谓的零形回指,就是在上下文语境中,动词的论元缺省了,没有以语音的形式出现,然而人们仍然可以通过上下文找回这个所指对象。

汉语是自由零回指语言,即零形式的出现或使用频率相当高。李和汤普森(Li & Thompson, 1979)在研究中发现:汉语中的零形回指远比英语广泛,并且汉语的零形回指不受语法结构限制,确定零形回指的所指对象,不仅需要语义和语用方面的知识,还需要语篇上下文信息。陈平(1984)认为,零形回指是汉语中非常普遍的一种语言现象,它几乎可以出现在句子中任一名词或代词出现的句法位置;零形回指在句中可以担任主语、直接宾语、间接宾语等句法成分。黄衍(1994: 257)指出:汉语中零形回指的出现非常自由,其使用的频率远远超过代词,回指语的先行语更多地由语用推理所决定,而非句法决定。李文丹(Li, W., 2004)也认为零形回指现象是汉语和英语的一大区别,在英语中广泛使用代词回指,零形回指的使用受语法限制,而汉语中代词回指的使用则少得多,零形式被大量使用。

在语言类型学中,零形回指属于代词脱落(pro-drop)现象中的一类表现形式。乔姆斯基创建的生成语法(Generative Grammar)理论提出的最具革命性的思想之一是:所有的语言在各自的不同特性之外都拥有一些基本的共性。在针对语言的特性和共性的研究中,代词脱落现象(在很多研究中被统一归类到空主语现象中)是被研究得最多、形式化最充分的语言特点之一。在很多语言中,处在主语位置上的代词可以不需要在句中显性表达出来,而以空语(null argument)形式出现,这种现象被称为空主语(null subject)或零主语(zero subject),也被称为代词脱落现象。赫姆伯格(Holmberg, 2010)提出将空主语语言划分为典型空主语语言(consistent NSL)、形式空主语语言(expletive NSL)、篇章型代词脱落语言(discourse pro-drop language)、部分空主语语言(partial NSL)。汉语被划分为篇章型代词脱落语言,又称为激进型代词脱落语言。这类语言的特点是篇章中的各种代词性成分都可以自由脱落并能从篇章上下文中获得还原。在篇章型代词脱落语言中存在一个允许激进的代词脱落的隐性的语法机制,在该机制中,上下文中可以被理解的先行语被复制到空代词的论元

位置上,从而使得脱落的代词可以获得还原。对于汉语中允许激进的代词脱落的这种隐性语法机制,虽然已有从句法、语义、语用和认知等多角度的分析,但尚未形成系统性的理论解释,也缺乏实证性的支持。对汉语代词脱落的深入研究有助于深化人们对于汉语的句法、语义、语用、篇章等各个层面的语言规律的认识,而通过和其他语言的对比研究,有助于语言共时特征的发现和解释、语言历史演变的研究、语言学理论的建设,以及单一语言研究的深化。

综上所述,汉语零形回指的认知机制的研究,一方面将在深层语言知识的层面为指代加工的认知机制提供更多的理论支持和规律破解,为中文信息处理中的指代消解研究提供语言学层面的理论支持;另一方面也会推动汉语代词脱落现象的语言类型学研究,在语言的共时和历时研究、对比语言学、对外汉语教学等领域也有重要的研究价值。

第二节 相关研究综述

西方的语言学者们从不同的角度对回指的本质和用法展开了研究,主要包括句法角度、语用角度、语义角度、功能角度、心理学角度、认知角度、计算语言学角度、篇章角度等。这些研究多数都是以英语为语料,由于英语中零形回指现象并不普遍,且用法较为规则,因此这些研究中对零形回指现象涉及较少。

相比之下,在汉语的回指研究中,零形回指受到了研究者们特别的关注,主要的研究内容包括以下几个方向。

1. 零形回指的认知策略和心理现实性

陶良(Tao, 1993; 2005)在解释人们在处理富含零形回指的汉语篇章时使用到的认知策略时提出了两个假设:①母语中的结构信息为人们对外语的加工提供了具体的认知策略;②这些认知策略又会对他们第二语言的学习产生重要的影响,并通过认知实验的方式对这两个假设进行了验证。陶良认为,汉语中零形回指的出现并非只由主题连续性所决定,而是受到语境的影响,因此语篇中的零形回指往往可以同时指向一个以上的候选先行语,至于先行语具体是哪个,则受到类似“出现指称”(Emergent Reference)等若干认知策略的影响。

杨宁(2008)试图通过引入话题因素(话题的连续与不连续)对汉

语篇中零形回指的消解进行系统的心理语言学研究,对汉语零形回指消解的心理现实性、时间进程和影响因素进行了初步探讨,并用认知实验的方式对自己提出的假设进行了验证。实验证明:在汉语篇章的阅读理解过程中,零形回指消解是一个复杂的思维和认知过程,但作为母语为汉语的人,我们头脑中积累了大量的图式和脚本,同时借助百科知识、语法结构、上下文关系和其他各种因素,我们可以在较短时间内判断出零形回指的所指对象。这说明汉语零形回指消解与代词回指消解一样,是一个快速和自动化的过程,无论是零形回指消解还是代词消解,其背后的心理机制应该都是符合人们认识世界的普遍认知机制的。

2. 汉语零形回指的使用条件

李和汤普森(1979)是较早关注汉语语篇中回指现象的学者。他们以《水浒传》和《儒林外史》为语料对汉语零形回指[他们称之为零代词(zero pronoun)]作了研究,发现零形回指是汉语的一个标准模式(norm),不需要任何显性的标记,而结构并非是控制零形回指的决定因素。他们从话语功能角度研究汉语中零形回指的选择,认为零形式出现于主题链(topic chain)中,其功能是将各分句连接成一条话题链;代词的出现标志着一条新话题链的开始,而名词则出现于一个段落的开头。零形回指使用的关键性因素是“连接性”(continuity),两个小句之间的连接性越高,就越倾向于在第二个小句中使用零形回指。所谓的“连接性”,是指一个小句与它前面一个小句形成一个语义单位的紧密程度,包含小句中各种句法和语义的特征,也包括说话者对语用情形的观察。

陈平(1984; 1987)是国内研究汉语回指现象较早和较有影响的学者。在他1984年的研究中,他将可预测性和可忽略性视为影响汉语第三人称零形回指使用的主要原因,因为可预测性和可忽略性具体地表明了篇章的话题连续性。其观点可以概括为:零形回指所指的可预测性和可忽略性越高,越倾向于选择零形回指,反之则相反。衡量可预测性的标准有三:①有效的或无效的竞争名词;②与前面小句连接性的高低;③可及性的高低。可预测性及其衡量标准之间的关系是:名词短语的有效性、连接性和可及性越高,则可预测性越高;反之,则可及性越低。1987年他发表的《汉语零形回指的话语分析》是国内第一篇系统地从话语的角度研究汉语零形回指的论文。该论文侧重分析了汉语话语结构特征对于零形回指的制约作用,提出以下观点:所指对象在话语的连续性是回指时使用零

形式的必要条件。连续性有微观和宏观之分。微观连续性取决于先行语和回指对象在各自句子的信息组织中的地位,体现在先行语的启后性和回指对象的承前性上。先行语在两种情况下启后性最强,一是做主语,二是作为新的信息成分出现在存现动词后面或者做普通动词的宾语。回指语做主语时承前性最强,做宾语时次之。先行语的启后性和回指对象的承前性越强,所指对象的微观连续性就越强。宏观连续性取决于先行语和回指对象各自所在的句子在话语组织中的关系。

蒲(Pu, 1997)认为,主题连贯性(thematic coherence)是零形回指出现和分布的决定性因素。主题连贯性越高,回指先行语的可及性就越高,因而使用零形回指的可能性就越大。当主题保持连贯时,即使在间隔多个小句的情形下,人们也会倾向于选择零形回指。当主题连贯性受到破坏时,则倾向于使用代词回指来取代零形回指。同时Pu也认为零形回指在汉语语篇中出现并不像之前一些文献中所认为的那么“自由”,而是受到语篇和语用原则限制的。

朱勘宇(2002)发现,在同话题推进的语篇中,零形回指的出现同前后相连的两个谓语小句的句法形式有关,将其称为句法驱动力。导致零形回指使用的句法驱动力包括两个因素:①相邻谓语小句的句法结构相似,称之为句法平行;②相邻谓语小句用关联词语连接起来。

屈承熹(2006)认为,话题位置上的零形回指,在话题链中标示为高连续性,这种连续性可以从既有话题展开,也可以从存现句导入的指涉对象展开。从既有话题展开的零形回指表明了当前话题的连续;从存现句导入的指涉对象展开的零形回指则表明当前话题发生转换。

曾立英(2008)从结构特征的角度研究了影响零形回指的因素,认为直接引语结构、叙事语结构、流水句结构、双名结构、存现结构、事件结构等都对零形回指的使用产生影响。

殷国光(2009)探讨了显著性及语篇中存在的话语标记对零形回指的制约,提出先行语所处的显著句法位置是制约零形回指选择的一个重要因素,同时具有衔接连贯作用的词语、句式、语境、背景知识、逻辑推理等作为广义的话语标记能够提示读者零形回指的先行语,因而具有上述标记的句子也倾向于采用零形回指。此外,层级结构也对零形回指的使用起到制约作用,但当零形回指与先行语各自所在小句之间存在插入语句(直接引语、评述语、插叙语句)时,这些插入语与零形回指或先行语所在小

句处于不同层次,篇章仍可使用零形回指。

3. 汉语零形回指的先行语解析

黄衍(1994: 149—150)基于新格莱斯原则(neo-Gricean principles)提出用数量原则(Quantity Principle)、信息量原则(Informativeness Principle)和方式原则(Manner Principle)分析回指的解决方案,并提出了“local subject > local object > matrix subject > others”(内嵌小句主语>内嵌小句宾语>主句主语>其他)的汉语零形回指先行语序列。

游(You, 1996)提出了分析零形回指的三个原则:最近原则(recency principle)、首要原则(primary principle)和分离原则(disjoincy principle)。最近原则的内容是:离下个小句的零形回指对象最近的那个成分,可能是零形回指的先行语。首要原则的内容是:在所有可能成为某个零形回指的所指成分中,“话题延续性”(topic continuity)的主要成分比其他成分更容易充当回指对象。分离原则的内容是:由“介绍/变化的世界”的谓语刚刚引进的新成分可能成为话题延续性的焦点。她认为,这三个原则在不单独依靠语用信息的条件下,可以找到某个零形回指的先行语。

李(Lee, 2002: 30—90)讨论了找回零形回指的先行语的三个原则:新近原则(Recency Principle)、开始原则(Opening Principle)和语境原则(Contextual Principle)。新近原则将零形回指语之前出现的最近的一个新实体选为零形回指的先行语。当话题从新实体向旧实体转换时,需要使用开始原则。而当需要更多的语境信息才能确认零形回指的先行语时,则需要使用语境原则。其中触发开始原则的因素包括结果标记、情绪表达、序数、关联词。她认为这些原则的操作是在话题延续性的层次结构上进行的。当某个新成分出现时,应考虑这个新成分和该成分之后的谓语之间的语义相符性。如果这两者不相符合,那么就在当前话题延续性之内一个个寻找零形主语的所指对象。

徐赳赳(2003)在《现代汉语篇章回指研究》一书中研究了名词、代词各种回指,也涉及了零形回指现象,提出基于语用和语义判定零形回指的先行语。他研究了汉语中多动词小句中的零形回指,认为语用和语义(主要是句子中动词的意义)是判别多动词句子中零形回指对象所指的主要因素,句法(这里主要指句子的语法结构)则处于次要地位。

毛(Mao, 2003)考察了《左传》叙事性语料中的零形回指。他认为话题连续性之外,其他一些因素,如固定的语用模式、对话、平行的句法

结构、连续转换所指等都对《左传》中零形回指的选择产生制约。

许余龙(2004)的《篇章回指的功能语用探索》一书从回指理解的角度研究了各种指称现象,其中也涉及了零形回指。许余龙认为,在篇章处理的某一刻,有两种篇章实体可以成为下一小句的期待主题:一种是前一小句的存现结构中那个无定名词短语引入的实体,另一种是前一小句主语/主题位置上的名词短语或该名词短语中的一个名词短语表达的实体。而前一小句动词宾语或副动词(coverb,或称介词)宾语位置上的那个名词短语表达的实体将是期待副主题。许余龙把汉语中的零形回指视为高可及性标示语。他提出了基于先行语的主题性和回指语的可及性的篇章回指确认机制。其中关于零形回指的确认原则如下:在其他条件相同的情况下,在语篇处理的某一刻,遇到的一个高可及性标示语表示:先行语是小句的主题,如果指称对象在小句的前半部分提及;或先行语是期待主题或显著的期待副主题,如果指称对象在前一小句中提及。

蒋平(2004a; 2004b; 2005)探讨了影响先行语可及性的因素。她认为除了阿里尔(1991)提出的指称距离、竞争度、凸显、一致性等因素外,句法位置、层次性和生命性也影响先行语的可及性。她分析了58689字语料中的2132例,从零形回指和先行语的句法位置以及话语结构层次的角度,以及先行语和回指词之间的关系和语义与语用的兼容匹配原则的角度分析了零形回指发生的条件,并提出了一些确认汉语零形回指先行语的解析原则。

王德亮(2004)根据汉语中的语法角色和主题的突出性,基于向心理论提出了汉语零形回指的解析算法,并用语料库实验进行验证。许宁云(2006)将向心理论全面而系统地应用于汉语语篇分析,推导出了一套用于解析和生成汉语语篇零形回指的计算模型,并进行了实证研究。段嫚娟等(2009)则在向心理论的语法角色显著性排序的基础上对汉语中语法角色的显著性进行了分级,并提出了六个基于向心理论的回指消解算法。

翁依琴(2006)以认知语法中的参照点模型(Reference Point Model)为理论基础,提出了凸显性阶列(Prominence Hierarchy)和宏观联系性(Macro-Connectivity)这两个概念,并提出了三条零形回指先行语的确认原则。凸显性阶列是由先行语和零形代词各自的位置所组成的。宏观联系性反映了领地之间的联系性,包括领地内和领地间的宏观联系性。零形回指只出现在联系比较紧密的情况下。

4. 汉语零形回指的信息化处理

以黄曾阳(1998)为首的 HNC [Hierarchical Network of Concepts (概念层次网络)] 团队对零形回指现象作了比较深入的探讨。HNC 理论认为句群中大句的组合形式主要有共享现象、照应现象和接应现象三种。其中共享现象为大句和小句之间出现语义块的整体或局部共享,是汉语中最常见、最重要的句群结构,零形回指主要发生在这类现象中。该理论对共享句的划分能够把汉语中零形回指现象以形式化的方法比较清楚地表达出来,因此在零形回指的信息化方面具有很好的应用前景。

宋柔(1992)分析了汉语叙述文中的小句前部省略现象,提出了该现象的树形结构(参照树),并分析了参照树的某些形式规律和语义规律,为计算机处理汉语篇章提供了形式化模型和算法设计的依据。他把研究对象的基本元素定位在标点句,把大部分的标点句主语零形回指问题看作跨标点句的主语成分共享问题,同时指出除了主谓关系外,定中、状中、述宾、述补、介宾、连谓等句法关系也会跨越标点句,因而提出了跨标点句的句法关系问题,并提出了研究这一问题的理论框架。同样进行跨标点句共享成分研究的还有张瑞朋(2007)。他在跨标点句句法关系的理论框架下展开了对跨标点句共享成分识别问题的研究。在标注语料和专项调查的基础上,分列大小一百多方面总结出跨标点句句法关系发生的各种约束条件。提出了跨标点句句法关系的形式化的约束规则,以便信息化处理。

此外曹军等(2001)提出了在层次结构的宏观控制下,利用谓词的语义进行分析来消解零代词的规则。马红妹等(2002)建立了汉语篇章上下文语境模型,制定了主语省略恢复规则,给出了主语省略恢复的算法。殷鸿等(2007)通过语义分析进行省略判断,通过从语义功能到语法功能的映射进行省略判定,采取优选统一句法位置上的 NP 结构来进行主语省略恢复的方法。侯敏、孙建军(2005)提出了针对机器翻译的零形回指解析算法,即先确认与标记零形回指,然后对位于主语、宾语位置上的零形回指进行处理。叶和陈(Yeh & Chen, 2007)提出基于词性标注和浅层语法而不是完全语法的方法来进行汉语零形主语回指消解。吴和梁(Wu & Liang, 2008)提出了基于实例推理方法的汉语零形主语回指消解。胡钦谱(Hu, 2008)分别采用向心理论算法和基于机器学习的决策树(decision tree)的算法对零形回指的自动解析进行了实验,结果表明决策

树算法比向心理论算法具有更好的解析力。

以上这四个方面的研究覆盖了心理学、语言学、人工智能等多个领域,对汉语零形回指的心理机制、语言规律、形式化消解等展开了全面的研究。在这四个研究方向中,汉语零形回指的使用条件和汉语零形回指的先行语解析是和语言学联系最紧密的一个方向。这两方面的研究虽然在研究重点上有所不同,但是本质上都是对汉语零形回指的认知机制的研究,因此在研究内容和研究方法上有很多交集。然而在对汉语零形回指的认知机制的揭示上,这些研究还存在以下的不足之处。

1. 研究的理论性和系统性不足

在从语言学角度对汉语零形回指的认知机制的研究中,有很多概念和规律过于笼统和抽象,缺乏形式化的表征。如李和汤普森(1979)所说的“连接性”,游(1996)、蒲(1997)和李(2002)都采用的“话题延续性”,这种“连接性”和“话题延续性”的本质是什么?如何判断两个相邻语段之间是否具有足够的“连接性”和“话题延续性”?陈平(1987)指出主语和新信息都具有“启后性”,在主语和新信息都具有强启后性的情况下,先行语如何选择?游(1996)提出的最近原则、首要原则和分离原则之间有何相互影响,如何交互作用?徐赳赳(2003)提出的判别多动词句子中零形回指对象所指的语用和语义因素具体包括哪些?许余龙(2004)提出的先行语的主题性如何判断?这些都需要从更具体的语言形式的层面对这些概念和规律进行系统化的描述。

在另一部分从语言学角度的研究中,虽然有较为具体的语言形式层面的规律描述,但是缺乏从心理学角度对这些形式化规律的理论性解释。如殷国光(2009)指出的显著性和层级结构对零形回指的制约有何理论依据?朱勘宇(2002)发现的导致零形回指使用的句法驱动力,屈承熹(2006)、曾立英(2008)发现的对零形回指的使用产生影响的汉语结构的形式化特征之下所隐含的心理认知机制是什么?以往的研究对这些问题都缺乏足够的解释力。

一些研究是以向心理论(如王德亮,2004;许宁云,2006;段嫚娟等,2009)、可及性理论(徐赳赳,2003;许余龙,2005)、认知参照点理论(翁依琴,2006)等西方语言学理论为基础,由于这些理论主要是以英语语料为基础创立的理论,且这些研究都是建立在单一的某一个语言学理论的基础上,因此当被用于汉语零形回指的研究时全面性不够,还需

要通过更多对汉语语言本身特点和规律的考察以及多种理论的结合才能形成更为全面准确的解析机制。

而基于人工智能和机器学习方法的零形回指信息化处理研究存在注重工程实践而忽视理论研究的问题,因此受制于弱语言知识,近年来在性能的继续提高上遇到了“瓶颈”,研究的焦点因此需要向深层语言知识,特别是结构化句法信息和语义信息方面的研究转移。

综上所述,在零形回指的认知机制上,虽然以往的研究涵盖了各个层面的因素,包括语法、语义、语用、语境、百科知识等,并发现了一些形式化特征,但是由于不同的研究者是从各自不同的层面展开的研究,因此这些认知因素是如何共同作用于汉语零形回指的?它们之间有何交互作用?这些形式化特征和各认知因素之间存在着怎样的对应关系?在这些研究中都未能提供足够详细和深入的解释。总的来说,以往的研究比较散乱零杂,有的侧重于心理机制研究,而对语言本身的特点和规律考察不足;有的侧重于语言规律的描述,而缺乏对这些规律背后的认知机制的研究,同时未能进行系统化整合;有的侧重于形式化解析方案的研究,在语言规律和认知机制的研究上受到了单一理论模型的限制,缺乏全面性。因此,到目前为止,对汉语零形回指的认知和解析机制尚未能形成完整的理论性解释和系统性描述。

2. 以语料库研究为主,缺乏对零形回指现象的认知机制的实验研究

除了陶(1993; 2005)和杨宁(2008)从心理学角度进行的研究外,绝大部分从语言学角度对汉语零形回指的实证性研究都采取的是语料库研究的方法。语料库研究的方法优点在于语料的真实性和自然性,可以对语言现象在语言中的总体分布特点获得较为全面的结论,缺点在于缺乏针对性和可控性,对语言现象背后的认知机制的解释力有限。因此以往的这些基于语料库研究的方法大多都只停留在对零形回指使用和先行语选择的规律的描述上,而没有办法证明这些规律的存在确实和汉语使用者对零形回指的认知策略相关。同时由于语料库规模的限制,对于一些零形回指的非典型规律缺乏足够的语料支持。所以无论是在研究的深度还是在广度上,语料库研究的方法都是存在很大的局限性的。

第三节 研究内容与研究方法

由于缺乏足够的理论性和系统性,之前针对汉语零形回指的研究对以此为试读,需要完整PDF请访问: www.ertongbook.com