

刘颖 著

儿童早期词汇句法 习得研究

ERTONG ZAOQI CIHUI JUFU
XIDE YANJIU



山东大学出版社

ERTONG ZAOQI CIHUI JUFU
XIDE YANJIU

策划编辑 张晓林

责任编辑 狄思宇

封面设计 张荔

ISBN 978-7-5607-5285-3



9 787560 752853 >

定价：36.00元

教育部人文社会科学研究青年基金项目(10YJC740074)
山东大学人文社会科学自主创新基金项目(IFW09091)成果

儿童早期词汇句法习得研究

刘 颖 著

山东大学出版社

图书在版编目(CIP)数据

儿童早期词汇句法习得研究/刘颖著. — 济南:
山东大学出版社, 2015. 5
ISBN 978-7-5607-5285-3

I. ①儿… II. ①刘… III. ①儿童语言—研究
IV. ①H003

中国版本图书馆 CIP 数据核字(2015)第 126424 号

责任策划: 张晓林
责任编辑: 狄思宇
封面设计: 张 荔

出版发行: 山东大学出版社

社 址 山东省济南市山大南路 20 号
邮 编 250100
电 话 市场部(0531)88364466

经 销: 山东省新华书店
印 刷: 山东和平商务有限公司
规 格: 720 毫米×1000 毫米 1/16
16.25 印张 308 千字
版 次: 2015 年 4 月第 1 版
印 次: 2015 年 4 月第 1 次印刷
定 价: 36.00 元

版权所有, 盗印必究

凡购本书, 如有缺页、倒页、脱页, 由本社营销部负责调换

用实证主义验证乔姆斯基假说的尝试

——序《儿童早期词汇句法习得研究》

盛玉麒

上个世纪发生的影响学术发展史的重大学术事件，莫过于“一个转向”、“两个革命”。一个转向指哲学的语言学转向；两个革命指“信息革命”和“语言学革命”。

哲学的语言学转向与 20 世纪的逻辑实证主义、科学主义、证明证伪等分析主义学术思潮密切相关。姑且不论其是否让西方哲学走出了本体论、认识论与真理观纠结的困境，但却的确将语言学这一古老的传统学科推向了备受瞩目的前沿和领先学科的地位。

以计算机和远程通讯技术相结合为主要特征的“信息革命”，实现了计算机从数据计算到信息处理的历史性跨越，从而揭开了人脑智能解放的序幕。“信息社会”已经成为国际化的流行语和随处可见的现实。我们每日每时都感同身受着信息革命带来的“大容量、高速度、快节奏”的便利和恩惠。

作为信息重要载体的语言不仅在信息革命中首当其冲，更由于乔姆斯基将唯理主义理论和形式主义方法相结合，刮起了一场语言学革命的飓风，以 1957 年出版的《句法结构》为标志，乔姆斯基相继问世的《句法理论要略》、《深层结构、表层结构和语义解释》、《支配和约束论集》等著作，在语言学界，哲学界、心理学界以及计算机信息处理领域都产生了巨大的影响。

支撑乔姆斯基“语言学革命”理论大厦的核心思想之一，是关于儿童语言能力先天遗传的假说。用直白的话说，就是“儿童可以说出他从未听过的句子”。

这种观点在笛卡尔理性主义知识获得论那里就有所阐述。实际上，不仅是笛卡尔，几乎所有为人父母者都会发现，婴儿从生下来到会张口说话只有短短一年半到两年的时间，而孩子在会说话之前就已经能够听懂很多话。这的确是奇迹般的事实。

以往的行为主义心理学用“刺激——反应”学说解释儿童获得语言能力的原因，伦内伯格提出著名的语言习得关键期假说。乔姆斯基的解释是把这一奇



迹归结为先天遗传。他认为,儿童之所以能够在有限的时间段内、有限的刺激条件下就能获得无限的远超于经验刺激范围的语言知识和能力,一定有一个先天植入大脑中的有关语言系统的解释规则。因为世界上有N种语言,因此,这个先天植入的语言能力应该是具有普遍适用性的,即“普遍语法”。出生后在什么语言环境中,该语言系统便激活大脑中的普遍语法,从而获得该语言能力。因此,乔姆斯基的整个理论大厦,就是通过内省重现人类大脑中先天植入的“普遍语法”。这样一来,不仅解释了为什么儿童会在很短的时间内掌握母语,也解释了为什么世界上不同民族的儿童都能“毫不费力”地掌握各不相同的母语。

这是一个基于假说的理论大厦,一个在假说基础上演绎推理出来的宏大体系。“语言能力先天遗传”是一个假说,“普遍语法”也是一个假说。“刺激匮乏”和“儿童可以说出他从未听过的句子”的论断算不上假说,只能算是一种推测,目前尚未进行穷尽式的实证主义的证明。

一切科学的结论都应该能够证明和证伪,应该经过实证主义路线的验证。所谓“大胆假设,小心求证”就是个中的普遍规则。乔姆斯基本人从“形式句法”到“标准理论”再到“管辖约束”直至“最简方案”,恰恰体现了探索创新“小心求证”的过程。

沿着“语言能力先天遗传”的假说进行演绎推理,是一种“求证”,通过儿童母语能力获得过程中的母语输入量的案例分析,也是一种“求证”,而且应该是更有意义的求证。

用真实的案例证明儿童所接受的语言输入到底是不是贫乏的?儿童的语言知识和能力能否从语言输入中获得?儿童又是怎样从环境语言输入中获得并有效利用这些相关信息的?这些都需要进行实证主义的“小心求证”。无论从复杂性和风险度来说,都是一个“大胆尝试”。

刘颖君的博士论文就做了这样的大胆尝试。

刘颖君是新华兄的硕士,具有扎实的汉语言文字学基础。毕业后留校在国际教育学院从事对外汉语教学。在职攻读语言学及应用语言学专业博士学位,尊师重道,聪敏勤奋,执着好学。

儿童语言研究领域有不少身兼父母及研究者的案例,其研究成果颇有价值,而乔姆斯基的先天语言能力假说一直未得到科学证实。攻读博士期间,刘颖君在初做母亲与兼顾学业的矛盾中选定了这样一个题目,也是机缘巧合。这也是我极力推荐并乐观其成的大胆尝试。因为乔姆斯基太伟大了,这个验证无论如何都是难得的贡献。本师殷焕先先生曾经说过,语言学的研究,贵在新理论、新方法、新材料。自己不能创造新理论,也要学习和借鉴新理论,总之要与新理论相关联。能与语言学革命的新理论相关联,的确是求之不得的灵感。

语料库语言学方法当时在汉语言文字学领域,还处于起步阶段,还在争论“语言学的语料库”还是“语料库的语言学”。能够采用语料库语言学方法,通过录音录像采集信息,转写自建儿童真实语境语料库,也应该是一个创新点。抽样建立婴儿“语前阶段”的语境信息语料库,进行词汇、短语、句法等统计分析,并与儿童语后话语进行比较研究,在许多方面都具有潜在的填补空白性质。

发展心理学告诉我们,3个月的胎儿就已经有了听觉感知能力。胎儿的大脑在3个月前是形成阶段,这是胎儿脑容量和脑细胞迅速增加的阶段。4~6个月,脑细胞数量 and 大脑皮层内部结构进一步发育,神经传导系统也开始发育。伴随听觉的发育就具有了信息感知、传导和记忆的功能。保守地说,5~6个月的胎儿就开始接受母语信息的“刺激”了。

因此,儿童语言能力获得的实证研究应该从这个时候开始。

这项研究说来容易,做起来却需要持之以恒的毅力和严谨扎实的学风。采用等距离按比例抽样采集信息的方法,然后要人工转写输入电脑,建立数字化文档。

记得有件让她很纠结的事情,就是孩子已经18个月了,还不会说话。当刘颖君满脸愁容地告诉我这个信息时,也着实让我感到吃惊。我们一起分析为什么,怎么办?突然想起,是不是家里有人说方言啊?一问,果然如此:爷爷、奶奶说方言,爸爸、妈妈说普通话。不同的语码对于孩子来说,要在大脑中建立不同的语言符号规则解释系统,自然格外费力了。已有多个案例证明家中说不同方言的婴儿,通常会比只说一种母语的儿童说话晚。从第20个月开始,刘颖君的孩子开口说话了,而且突飞猛进,这让所有人都兴奋不已。语料库的材料日渐充实。孩子的话语素材中词汇、句法现象,丰富多彩。描写、分析、解释等研究自然步上了快车道。

刘颖君的研究结果发现,成人话语输入对儿童语言习得影响作用明显,输入频率高低与儿童习得快慢呈正相关。后天语言环境通过高频率、适合交谈情景和儿童需求的、略前于儿童语言发展水平的刺激,与儿童的认知能力相互作用相互促进,共同推动儿童语言的发展。

刘颖君的研究对于验证乔姆斯基的“语言能力遗传”假说,只是一个个案的尝试,但是,在理论和实践上的贡献却是显而易见的。因为,正如鲁迅说过的比喻那样:“其实地上本没有路,走的人多了,也便成了路。”论文完成后,我知道刘颖君仍然坚持对孩子语言发展的实况定时抽样录音录像。现在的书稿,所涉语料在博士论文0~2岁半的基础上,延长到了5岁。她的研究课题先后获得山东大学自主创新基金和教育部人文社科青年基金立项支持。作为该课题的代表性成果,曾在《语言教学与研究》2014年第5期上发表论文《环境语言输入与



儿童语言习得相关性计量研究》。

把录音语料转写建库是一项细致的苦差事,从语料中发现有研究价值的内容更是复杂繁琐的工作。反复梳理几万字的语料最后提取出来可以呈现给读者的结论有时只有两三千字,甚至只是一张表。

基于语料库的儿童个案母语习得研究,是当代语言学研究的一个重要领域和方向,甚至是重要的新兴交叉学科。这种真实文本语料库对于语言学、心理学、认知科学、计算机信息处理乃至逻辑学、哲学等,都具有非常重要的、无法替代的学术价值和资源价值。相信一定会有越来越多的学者关注这个领域,参与这个研究。

刘颖君的博士论文即将由山东大学出版社出版,这是她前期学术研究的一个小结,希望成为该领域研究中承前启后的一个亮点。博士选题具有可持续研究的价值,成为值得终身研究的事业,的确是一件值得祝贺的事情。机遇和挑战并存,奉献与成功相携,寄望于刘颖君及学界同仁,来日方长,任重道远。是为序。

2015年4月12日
于山东大学老校三宿舍

目 录

第 1 章 绪论	(1)
1.1 释题	(1)
1.2 本研究的历史与现状	(3)
1.3 本研究的目标、内容、思路与方法	(10)
1.4 本章小结	(15)
第 2 章 语料库建设	(16)
2.1 语料采集	(16)
2.2 录音转写	(19)
2.3 文本处理	(22)
2.4 分词校对	(26)
2.5 本章小结	(30)
第 3 章 语前阶段的成人话语	(31)
3.1 词	(33)
3.2 短语	(39)
3.3 句法结构	(40)
3.4 本章小结	(42)
第 4 章 成人与儿童语料的计量分析	(43)
4.1 词句统计	(43)
4.2 输入与输出的相关性	(50)
4.3 本章小结	(60)
第 5 章 儿童的早期理解状况	(61)
5.1 儿童对体态语的理解	(61)
5.2 儿童对成人话语的理解	(64)
5.3 儿童话语理解状况(一)	(76)
5.4 儿童话语理解状况(二)	(97)
5.5 本章小结	(122)



第 6 章 儿童的表达状况	(124)
6.1 非言语表达	(124)
6.2 最早 50 个“词”	(133)
6.3 领属结构的表达	(141)
6.4 问句的表达	(145)
6.5 “把”字句的表达	(157)
6.6 本章小结	(166)
第 7 章 结 语	(168)
7.1 研究结论	(168)
7.2 创新性	(169)
7.3 研究价值及意义	(170)
附 录	(173)
附录 1 成人与儿童名词对照表	(173)
附录 2 成人与儿童动词对照表	(181)
附录 3 成人与儿童副词对照表	(189)
附录 4 成人与儿童量词对照表	(191)
附录 5 成人与儿童代词对照表	(192)
附录 6 成人与儿童拟声词对照表	(194)
附录 7 成人与儿童方位词对照表	(196)
附录 8 成人与儿童数词对照表	(197)
附录 9 成人与儿童叹词对照表	(198)
附录 10 成人与儿童介词对照表	(199)
附录 11 成人与儿童语气词对照表	(200)
附录 12 成人与儿童助词对照表	(201)
附录 13 成人与儿童连词对照表	(201)
附录 14 成人与儿童区别词对照表	(202)
附录 15 儿童词首次听到与首次说出时间对照表	(203)
附录 16 31~60 个月儿童与成人“把”字句原始语料	(228)
参考文献	(240)
后 记	(248)

第1章 绪论

1.1 释题

1.1.1 研究缘起

乔姆斯基在儿童语言习得问题上始终坚持内在天赋论,刺激贫乏论(the poverty of stimulus argument)是支持这种理论的最强有力的依据,所谓刺激贫乏是指儿童所接受的语言输入和最终获得的语言能力之间是不对称的,这种不对称使生成语言学家相信语言习得的成功得益于天生的普遍语法(UG),正是与生俱来的普遍语法引导儿童克服语言输入的缺陷,顺利掌握语言。儿童习得语言的过程就是受经验触发为普遍语法赋值,从而发展到个别语法的过程。乔姆斯基将普遍语法定义为儿童通过生物进化和遗传获得的存在于一切人类语言之中并为人类先天具有的知识,同时将后天语言环境所起的作用降低到最低限度,即后天语言环境只起触发语言习得机制(LAD)的作用。然而普遍语法究竟包含哪些内容?儿童在语言习得过程中是怎样为普遍语法赋值的?语言习得机制是如何运作的?这些问题生成语法学家都未能做出回答,因此乔姆斯基的语言习得理论只是一种推论,这种推论是建立在刺激贫乏论基础上的。如果刺激贫乏论不能成立,将从根本上动摇语言天赋论的根基,从而也会动摇生成语法的根基。那么儿童所接受的语言输入到底是不是贫乏的?这些语言输入中是否包含了儿童习得语言所需要的相关信息?儿童的语言能力是否能从语言输入中获得?这些正是本书要探讨的问题,为此我们选取儿童早期(5岁以前)语言发展作为研究对象,在细致描写该时期儿童所处的语言环境及其理解与表达状况的基础上,将儿童的理解与表达、儿童的语言输出与所接受的输入加以对照,从中寻找问题的答案。

语言获得的可能性是儿童语言研究不可回避的课题,也是语言的本质问题之一,以往的儿童汉语研究大多关注儿童在哪个年龄段说出了哪些话,即儿童在某个年龄段的语言表达状况,而较少关注儿童在某个年龄段的理解状况、儿



童所接受的环境语言输入状况以及环境语言输入在儿童语言习得中的作用。理解先于表达,也是表达前不可逾越的重要阶段,理解属于习得范畴,研究环境语言输入状况对探索儿童如何习得语言也有重要意义。如果儿童能说出某个语言表达式,那么我们可以追溯环境语言中是否存在该表达式的输入证据,即进行输入和输出的对比研究,还可以追溯儿童能说出的表达式是否包含在儿童已经理解的环境语言输入中,即进行理解与表达的对比研究。基于以上认识,本课题在大规模语料库基础上展开对儿童早期词汇、句法的习得研究,从理解与表达、输入与输出的对比及其相关性研究中探讨环境语言输入对儿童语言习得的影响和作用,进而探讨儿童语言获得的可能性问题和儿童语言习得机制。

儿童早期语言研究因为研究对象的年龄较小,其认知和语言表达尚未达到成人水平,尤其是在儿童会说话之前。考察儿童对成人话语的理解状况只能通过一些外围可见因素如体态、声音或简单语音等来推测判断,而这些又有很大的个体差异,即便是进入语言阶段的儿童,其语言表达正处于发展与完善过程中,因此不熟悉被观察儿童的研究者很难仅仅凭录音录像资料得出切合事实的结论。本研究的观察对象为笔者的女儿,笔者与其朝夕相处,得以最大限度地近距离观察、记录其语言的发生与发展情况,为本研究提供了可靠的语料来源。

以上所述,构成了本研究的最初动力。

1.1.2 术语使用

(1)关于习得

儿童获得母语的语言能力是在日常生活中自然进行的,其中也有学习方式,尤其是在当今社会重视早期教育的大形势下,父母会利用幼儿读物、识字卡、幼教片等引导儿童进行有意识地学习,部分城市孩子还会接受早教中心的课堂教育^①,但这些有意识的学习活动并非儿童习得母语的主要方式,与成人在课堂上的学习方式也有较大差别,因此儿童掌握母语可以笼统称之为习得。

从以往文献看,“习得”一词更多地用于儿童言语表达,但实际上儿童习得母语的过程包括两方面,话语理解和言语表达,而且从发生顺序上说,理解先于表达,理解为表达奠定基础。本文从儿童理解和表达两方面探讨儿童语言习得机制,除特别说明外,“习得”一词包括这两方面内容。

(2)关于时间表示法

本文研究的是个案儿童(下文简称A)5岁前的语言发展,为行文简便,且考

^① 据我们了解,目前国内城市部分3岁以前接受早教的孩子一般每周去一次早教中心,一次课50分钟左右,但几乎没有一个孩子能够做到每周都去,因为孩子生病、天气不好、家长有事等原因,常常请假。



虑低龄儿童语言发展迅速多变、间隔时间短的特点,儿童年龄用月龄或“月龄:天数”的形式来表示。如:28个月,9:20(9个月零20天),37:02(37个月零2天)。

1.2 本研究的历史与现状

1.2.1 国内外的儿童语言研究

1.2.1.1 国外的儿童语言研究

人类对儿童语言的兴趣可以追溯到古埃及,毫无疑问,母语习得是所有语言习得中最成功的一种方式,语言习得理论的研究就是从儿童第一语言习得的研究开始的。心理学家、语言学家和语言教育家希望通过揭开儿童怎样习得第一语言之谜,了解人的语言能力是如何获得的。

国外的儿童语言研究大致经历了三个阶段:早期的儿童语言研究是作为心理学的一部分进行的,大多是一些心理学家用日记法收集自己子女的语言发展资料,为揭示儿童的心理发展和测试儿童的智力服务,这一时期的研究也比较粗糙。第二阶段是在行为主义影响下的研究,认为语言习惯是在一系列刺激和反应中学习得到的。第三阶段是在心理语言学诞生之后的研究,这一阶段的研究受皮亚杰和乔姆斯基的影响,皮亚杰在认知发展理论的基础上提出了儿童语言发展的相互作用论,乔姆斯基的学说促使人们开始思考儿童对语言知道些什么。

随着认知科学的兴起,一种关于语言和人类语言能力的新观点——认知功能语言学日益受到关注,该理论重视语言的符号维度,强调语言结构是在具体运用过程中形成的,它对语言获得理论带来极大启示,研究者开始深入探讨语言功能、社会文化环境以及儿童的主动建构在语言获得中的作用,Michael Tomasello 提出的用法建构语言获得理论代表了这种新的研究取向。

时至今日,儿童语言获得到底是先天决定的还是后天作用的,仍在争论中。其实不管是先天决定论、后天环境论还是先天后天相互作用论,都承认儿童习得语言过程中既有先天因素又有后天因素,只不过对于哪些因素是先天的,哪些因素是后天的,先天与后天在习得过程中的作用大小以及如何起作用的争论不休,因此弄清这些问题是今后儿童语言研究的方向和重点。

1.2.1.2 国内的儿童语言研究

国内的儿童语言研究起步较晚,始于五四运动前后,医学界、心理学界与语



言学界都对儿童语言有大量研究,但研究目的、研究重点各有不同。

医学界的儿童语言研究主要集中在生物学基础研究、语言发育进程研究、语言发育异常标准的制定和应用三个方面。生物学基础研究又分为发音器官的进化、大脑功能的分化、语言基因的发现三个方面;语言发育进程研究分为前语言阶段(Pre-language Stage)和语言阶段,并以语言阶段为研究重点;语言发育异常标准的制定和应用是医学界的重要任务,他们通过收集儿童语言发育进程资料来制定汉语语言发育筛查和诊断量表,制定预防和治疗语言障碍的干预措施。但医学界常用问卷调查或医务人员询问儿童家长完成量表填写的方式收集正常儿童各年龄段语言发展数据,以此作为拟定儿童语言发育迟缓或可疑迟缓的筛查标准,这种收集方式的准确性、可靠性不够。

心理学界的儿童语言研究是为探究儿童心理与行为发展服务的,记忆、思维和语言构成了儿童心理发展的三大领域之一——认知领域。五四时期我国引进了许多西方儿童心理学的研究成果,同时也进行了最初的儿童语言研究,有个案追踪记录,如陈鹤琴对其儿子进行的两年半的语言发展记录,也有对西方成果的检验,如黄翼、祝雨人进行的儿童语言的社会功用研究,补充或修正了皮亚杰的结论。1978年,以朱智贤为首的北京师范大学心理学系研究小组和以朱曼殊为首的华东师范大学心理学系研究小组开始对儿童语言进行长期调查、实验和研究,取得了丰硕成果。北京大学心理系的许政援等曾对多名尤其是0~3岁儿童的语言发展进行长期仔细的追踪研究,并据此提出一些有说服力的论断。^①近年来,华东师范大学儿童语言研究中心利用国际儿童语言资源交换系统 CHILDES(Child Language Data Exchange System),开展了有关汉语儿童语言发展和教育的研究,尤其在儿童语言运用能力、语言交流行为方面有所进展,逐步与国际接轨。^②

以语言本身作为研究对象、试图揭示人类语言本质以及儿童如何获得语言能力的儿童语言研究是在语言学界展开的。20世纪80年代中期在一些心理语言学著作中开始较多地介绍国外儿童语言研究的成果和方法,如桂诗春的《心理语言学》和《实用心理语言学纲要》。90年代出现了一些开拓性研究成果,如李宇明、唐志东、陈前瑞等采用个案与群案观察相结合的方法研究儿童问句系统的习得,创造性地运用了多种儿童语言测试方法,总结了儿童问句系统的习得规律和相关理论问题,丰富了儿童语言研究的一般性原则与理论。

① 参见李宇明、汪国胜《中国儿童语言研究概述》,《语言学通讯》1988年第3、4期;眸子《国内儿童语言研究鸟瞰》,《中文自学指导》1991年第4期;许政援《三岁前儿童语言发展的研究和有关的理论问题》,《心理发展与教育》1996年第3期。

② 详情请参见网站:<http://www.childes.ecnu.edu.cn> 和 <http://childes.talkbank.org>。



李宇明《儿童语言的发展》是立足语言学领域的一部儿童语言学著作,其目的就在于研究儿童语言发展本身,该书充分利用汉族儿童语言发展的材料,比较全面、细致地刻画了儿童语言发展过程中的语音、词汇—语义、语法、语用等方面的现象和问题,并紧紧围绕语言学这一中心,引介述评儿童语言研究历史、儿童语言发展基本理论,探讨儿童语言发展的阶段性和顺序性等,从实践上和理论上丰富和发展了儿童语言学。

同一时期的孔令达、周国光、李向农、王葆华等采用横向调查法、实验法研究儿童实词、虚词、句式等的习得情况,在充分描写的基础上,对语言习得和语言发展理论的重要问题进行了深入探讨,如从语言结构能力发展的角度重新定义儿童语言发展的三个阶段:词语法阶段—词组语法阶段—句语法阶段,并提出儿童习得基本句法结构的主要机制是句法分解和句法合成,最后初步建立了汉语语言习得理论体系。

此外还有李行德等进行的汉语儿童早期习得项目(Chinese Early Child Language, CELA),以及张云秋等在汉语儿童副词习得、儿童语言获得的动因和机制等方面的研究。

近十年来的儿童语言研究呈现出多元化的特点。除了以传统方式继续探讨儿童语言在语音、词汇、语义、语法、语用等语言基本方面的习得、发展情况外,一些学者意识到建立统一格式的儿童口语语料库是增强儿童语言研究科学性的重要手段,因为仅供个人或在有限范围内使用的语料,对于语料资源来说是极大的浪费,同时其他研究者无法对数据及结论的真伪进行鉴别,因此一些研究者试图把自己掌握的儿童语料以计算机能够识别、处理的方式保存起来,表 1-1 是目前国内外现有汉族儿童口语语料库基本状况。

表中有些语料库采用了 CHAT(Codes for the Human Analysis of Transcript)录写系统,该系统是一套复杂但又十分灵活的多级赋码方案,专为计算机录写自然话语而设计。赋码内容包括基本的话语文本类型、详尽的发音信息和话语的句法分析信息。正确的 CHAT 文本格式如下:

- (1)用英语进行正常赋码时,文档中的每个符号都必须是 ASCII 符号。
- (2)每行必须以硬回车结束。
- (3)每行的开头必须是@Begin。
- (4)每行的结束必须是@End。

(5)必须有@Participants 行。内容为代表参加者的三字码,如:参加者的名字、参加者的角色,此行必须紧跟@Begin 行。

(6)以“*”开始的行表明实际说过的话。这些被称为“主行”,每一个主行都只能记录一句话语。当说话者一次说几句话时,必须分行记录。



(7)在主行的“*”之后的三个字母必须大写,然后是冒号和说话内容。

(8)以%号开始的行表示录写者或研究者对背景的注释。通常这些行包括赋码和注释,称为“附属行”。

(9)以%号开始的附属行后是三个小写字母。如“phonology”中的“pho”,一个冒号和注释。

(10)摘要行后接主行和附属行。

另外 CHAT 录写系统还规定了详尽的赋码方式、原则、操作步骤等,十分全面,经过赋码后的 CHAT 文本可以运用 CLAN(Computerized Language Analysis)程序进行自动分析。李行德等进行的 CELA 项目就采用了该录写系统,而且所用语料已成为 CHILDES 国际儿童口语语料库的一部分。

表 1-1 现有汉族儿童口语语料库概览^①

语料库名称	研制者	年龄段	儿童人数	语料格式	语料规模	采集背景
CanCorp	香港中文大学 李行德等	1;5—2;8	8	录音	171 个 14 兆	自然状态
HKU	香港中文大学 Zehava Weizman Emily Ma	2;6—5;6	70	录音	128+70 个 28 兆	自然状态
Beijing	香港中文中学 Tardif(夏玲)	1;9—2;2	10	录音	50 个 3 兆	自然状态
Context	香港中文大学 Tardif(夏玲)	2 岁	25	录音	12 个 0.5 兆	自然状态
Chang	台湾师范大学 Cienju Chang	4 岁和 6 岁	24	录像	24 个 0.7 兆	特殊场景
Taicorp	台湾中正大学 蔡素娟	1;0—3;0	14	录音	230 万字	自然状态
Zhou	华东师范大学 周 兢	1;0—4;0	15	录像	52 个 0.6 兆	特殊场景

^① 参见嵇红英《儿童语料库建设与幼儿词汇语义分类》,南京师范大学硕士学位论文,2007 年,第 5 页。



续表

语料库名称	研制者	年龄段	儿童人数	语料格式	语料规模	采集背景
现代汉语口语 语料库子库	中国社科院 沈家煊主持	?	?	MD	?	?

说明:语料规模中“171个”表示171个CHAT^①文件,采集背景中的“特殊场景”是指由研究者设定的背景,所采集到的语料是在研究者设定的背景下儿童所产生的语言。“?”表示至今仍未收集到相关资料。

应用CHAT录写口语语料工作量极大,且低龄儿童的发音不够清晰准确,需要熟悉其说话特点的家人仔细辨认确定,再加上研究目的和抽取的数据类型不同,本研究未采用该录写系统。但本文在建设语料库时,充分借鉴了CHAT系统的赋码原则及方式,制定出一套适合本研究的录写及数据处理方法,如根据研究目的需要将原始数据分解到最小单位、添加必要的检索、控制信息等,这些都保证了后期研究工作的顺利进行。

1.2.2 汉语儿童早期语言研究

根据语言出现与否,可把儿童早期语言研究划分为语前和语后两个阶段,语前阶段是前语言阶段(Pre-language Stage 或 Prelinguistic)的简称,即语言出现前的阶段,也就是儿童开口说话之前的阶段,语后阶段即语言出现之后的阶段,也就是儿童开始使用并逐渐掌握语言工具与他人交际的阶段。语前阶段与语后阶段的划分,以语言的出现为标志,但实际情况是,很难断定儿童在哪一刻达到了真正语言阶段,而且也很难肯定单字的真正语言价值。^②有一点可以肯定,语言的出现必定以前期的话语输入和儿童自身的认知发展为基础,从理论上来说,可以这样定义语言的出现:

a. 说出能指与所指相结合的符号,即音义结合的符号;

b. 该符号在其母语系统中具有可识辨性,即与其母语中的非叹词非语气词的语言单位大体相合。(因为叹词和语气词主要是标音性质的,这类词的音、义、音义的结合都比较模糊,不好作为判断语言是否出现的标准)。^③

^① CHAT(Codes for the Human Analysis of Transcripts)是国际儿童口语语料库 CHILDES 的录写系统。

^② 参见靳洪刚《语言获得理论研究》,中国社会科学出版社1997年版,第42~43页。

^③ 参见李宇明《独词句阶段的语言特点》,《中国语言学报》总第7期,语文出版社1995年。