

论现代汉语语法研究中的证据问题^{*}

陈 平

(澳大利亚昆士兰大学语言与文化学院)

提 要 语法研究所用经验证据有三个主要来源:自拟、真实出现的语言材料,包括口语和书面语,以及实验或调查。现代汉语语法研究始于黎锦熙 1924 出版的《新著国语文法》。1950 年代之前的现代汉语语法著作,根据的几乎全部是书面语材料,包括早期白话文和现代白话文作品,并且大都标明出处,以示有书为证。这种做法延续到 1950 年代。从 1955—1956 年公布的普通话定义也可以看出,它代表了当时的共识。1960 年代之后,尤其是近三四十年来,通过自拟获取例证的做法渐渐通行开来,更为重视口语,尤其是北京口语的研究,从书面语直接引用或改写的例句也可以不必注明出处。这些变化反映汉语语法研究者对自己语感的自信在不断增强,反映现代汉语语法学界对于什么是合语法或不合法语的用法有了越来越多的共识,也反映现代汉语标准化程度在不断提高。现代汉语语法研究应该如何获取证据,涉及许多同现代汉语标准语的起源与发展有关的问题,主要围绕如何看待作为语料的现代汉语标准语、口语和书面语这三者之间的关系,如何把握自拟、利用真实语料以及实验/调查这三种获取证据的方式各自的特点。本文的主要目的,是梳理有关问题的来龙去脉,并指出某些问题的症结所在。

关键词 现代汉语 语法 普通话 北京话 口语 书面语 白话文

DOI:10.16027/j.cnki.cn31-2043/h.2018.05.001

一、语言研究中理论依据与经验证据的区别及其来源

任何一个论点,首先需要有可靠的论据,然后经过严密论证才能成立。无征不信,说话要有根据,不仅是日常生活中的基本思维方法和行为准则,更是科学研究中的铁律。论点、论据和论证三者中,论据是基础,论证是方法。基础不牢靠,论证和论点都谈不上。胡适常将“有几分证据说几分话,有七分证据不能说八分话”写成条幅送人。言必有据,体现了五四时代提倡的科学精神,也是包括乾嘉学派在内的中国优良学术传统的重要内容。无论是传统语言学还是现代语言学研究,论据最常见的表现方法就是例证,用例子说明问题。黎锦熙早年提出“例不十,不立法”,更是为语言规律能否成立给出量化标准。王力又将这句话引申到反例上

^{*} 2017 年 12 月,复旦大学举办中文学科百年庆典,作者应邀以“现代汉语的起源:再论‘白话文运动’”为主题作学术报告,本文根据报告记录改写。对祝克懿教授的邀请和陈光磊教授及其他听众提出的宝贵意见,作者表示深切的感谢。

去提出“例外不十，法不破”。两位学者主张的是同一条原则：寻绎语言规律的过程中，例证必不可少，而孤证或者出现次数很少的例子不足为凭。

理论立场 vs 经验证据

强调例证的同时，我们也要避免陷入“唯证据论”的误区。语言学既是经验科学，也是理论科学。涉及具体研究课题时，理论问题和实证问题往往密切相关，甚至缠绕难解。在讨论某个论点的过程中，我们或许会认为，论点能否成立完全取决于能否找到证实或证伪的例证。这种态度是片面的。面对我们研究的问题，我们首先要区分，这是理论问题，还是实证问题？哪些方面是理论问题？哪些方面是实证问题？大多数理论观点应该通过经验证据来检验能否成立，但也有许多理论观点，尤其涉及定义问题时，经验证据的多寡强弱往往对论点能否成立并无实质影响，甚至毫无影响。划清经验证据的效用边界，能使我们对例证在语言学研究中所能起到的作用，有更清醒的认识。我们下面以1950年代汉语语法学界三次大讨论为例，说明这个道理。

首先是汉语词类划分，这是至今都充满争议的研究课题。主要有两个相连的问题，一是汉语有无词类，或者说能否划分出名动形这样的词类；二是如果认为汉语有词类，那么划分词类的标准是什么？如何划分词类？这个问题之所以聚讼纷纭，一定程度上是因为涉及的问题，有些是理论立场问题，有些是经验证据问题。高名凯主张，区别词类，只能以狭义的、标明各种词类的特殊词形变化为标准。这个主张有两方面的内容，一是经验证据问题：汉语中能否找到他说的狭义的特殊词形变化。如果找得到，那汉语能分出词类；如果找不到，就无法为汉语划分词类；二是理论立场问题，高名凯认定，只有他说的特殊词形变化才是划分词类的标准，其他如意义、功能、分布等都与此无关。这是典型的由定义决定的理论立场，不涉经验证据。评判这类理论立场，主要根据其逻辑是否自洽、周延，对有关语言现象的说明是否全面、透彻，描写和解释的方式是否简练、深刻，是否导向有重大意义的研究课题等标准来决定它们的优劣。在高名凯词类划分标准问题上，就理论立场而言，经验事实本身不足用作证据判定他这个主张的对错，因为旁人用来证明汉语可以分出词类的语言事实，用高名凯的定义来衡量，作为划分汉语词类的标准，都不合格，因此没有充任证据的资格。黎锦熙（1924）主张，“凡词，依句辨品，离句无品”。这个观点也有两个部分，第一部分“依句辨品”，是他主张用来辨别词类的方法，是个理论立场问题，无法用经验事实对它证实或证伪，“离句无品”则是个经验事实问题，如果有语言事实可以证明，离开句子，照样能对汉语中的词进行分类，那就是用证据推翻了“离句无品”的论点。其他研究者给出“离句无品”的反证之后，黎锦熙改变了自己的观点。他在《新著国语法》1951年“今序”中说：“‘凡词，依句辨品’是对的，但又说‘离句无品’则是不正确的。”（黎锦熙1924/1992：25）其实，“依句辨品”是理论立场问题，不受经验事实左右，只有优劣之别，没有对错之分。讨论汉语词类划分的文献中，涉及有无所谓兼类词，以及兼类词的划类策略等问题时，一般会用到“常体”“变体”等概念，取决于具体用法在语言中使用频率的高低，这是经验事实问题，可以通过统计数字来判定。

主语宾语的辨识在很大程度上也是一个理论立场问题，无法根据经验事实决定对错。汉语传统语法著作中，往往将施事认作主语，受事认作宾语，施事受事是初始成分，主语宾语是次生成分，后者的认定取决于前者。“他卖了栋房子”是主谓宾句，“房子他卖了”成了宾语前置句。美国结构主义理论框架中，主语宾语也都是句子的次生成分，由句子的结构关系

(constituency), 主要是层次关系和前后语序决定, 可由短语结构语法定义。句子 S 下辖两个直接成分, NP 和 VP, NP 就是主语, VP 是谓语。VP 下辖 V 和 NP, V 后面的 NP 就是宾语。主宾不在同一个层次。在一些以依存关系 (dependency) 为主的理论模式中, 主语宾语同为初始成分, 在句中位于同一层次, 一般不设谓语的概念。在这几大类理论模式中, 主语宾语的辨识由理论立场决定, 同经验事实没有多大关系。在功能语法研究传统中, 往往将主语看作施事、话题等语义、语用成分的语法化结果, 具备某些典型的语法行为特征, 涉及许多经验事实范畴里的问题。

至于单复句问题, 几乎完全取决于研究者自己设定的标准, 是个理论立场问题, 旁人无法根据经验事实来判断对错。所谓单复句之争, 因为各人所用标准不同, 往往实际上并没有许多实质内容。

语法研究中经验证据的来源

语法研究所用的例证有三种主要来源。首先是自拟, 研究者根据自己的语感, 就有关语言现象自造例句, 并以此为据, 对有关论点加以论证。自拟例证可靠程度有多高, 主要取决于本人对于该语言的熟悉程度。一般说来, 研究者只有对自己的母语才有资格通过自拟获得可靠的例证。

第二种来源是真实出现的语言材料, 可以是口语, 也可以是书面语。录音机出现之前则基本都从书面语获得例证, 即所谓的“书证”。研究古代语言, 因为不能起古人于地下, 书证向来是唯一选择。从语料价值来看, 书证其实有很大缺陷。首先, 书证表现的主要是书面语而不是口语, 而就语言的本质属性来说, 口语是第一位的, 书面语是口语的文字表现。书面语同口语之间一般来说总是有距离的, 书面语作者往往时间充裕, 在遣词造句方面可以仔细推敲, 倾向使用比较正式的词语; 而口语则相对而言比较随意, 在用词和组织条理上有时不是太齐整。即使尽可能忠实再现口语内容的文字记录, 也无可避免地会丢失许多重要信息, 如说话人语音语调上的快慢轻重等。二十世纪录音机的普遍使用, 对语言学家来说, 无疑为口语研究带来很大方便。从事社会语言学、话语分析等研究工作的人经常使用录音材料作为研究对象。过去几十年来, 出现了汇集数亿甚至数十亿字词的书面语和口语语料库, 为研究者提供了可用电脑方便地搜索、提取、归类、对比的语言材料, 大大丰富了语言研究经验证据的种类和数量。

值得强调指出的是, 利用真实语料搜集证据的方法, 还有一个严重的缺陷。当代语言学在研究方法上有别于传统语言研究的创新之一, 是大量利用不合语法或不适用所在语境的句子, 从反面揭示相关规则。这种例证可以通过自拟或实验/调查方法获取, 但在实际出现的语料中是见不到的。

第三种是通过实验或调查方式获取例证。实验或调查采取的形式可以多种多样, 规模可大可小。有的是按照精心设计的方案请多人参与的正式实验和调查, 也有的只是向个别人简单地问几个问题。现代汉语语法研究领域里通过大规模实验和调查获取证据的情况并不常见, 但研究人员就某个语言成分的用法咨询周围说母语的人的意见, 则是非常普遍的做法。

研究者的理论背景、研究领域和课题, 很大程度上决定了他们主要使用哪一种方式获取证据。致力构建形式理论模式的生成语言学派几乎完全使用自拟的例证, 认为对于语句是否符合语法的判断, 是说话人语言能力 (competence) 的一个核心内容。心理语言学、社会语言学、语言习得、方言研究等所用的经验证据, 基本上都是通过实验或调查取得。话语分析, 以及最

近二三十年发展很快的基于用法的一些理论学派,主要用的是实际出现的真实语料,尤其是口语材料,而语言成分所在的语境则是语料不可分割的一部分。功能语言学、认知语言学背景的语言学家,一般根据研究课题和研究目的,在自拟、利用真实语料与实验/调查三种方式中间,选择最方便、最有效的方式获取证据,常常采取一种以上的方式。

获取例证的三种主要方式各有特点、各有利弊,但都有一个可靠性问题。语法研究中,某个句子是否能说、语义应该怎样理解、用在某个特定的语境是否适当等,都涉及研究者的判断问题。自拟例证完全是研究者基于语感的个人判断,主观性最强。真实出现的语料以及实验和调查取得的结果,并非能自动成为可靠的证据。对于前者,研究者从中挑选可以用作证据的材料,选择的过程,就是个人主观判断的过程。虽然实验和调查结果相对而言客观性较强,但实验和调查方式的设计,以及所得结果的解读,许多时候也是个人判断的结果。个人基于语感的判断是否可靠,是语言学界许多争议的症结所在。

三种方式中,自拟无疑最省事省力,是语法研究中最常用的举证方式。但是,研究人员虽然讲的是同一语言,语感往往可能因国别、方言、年龄、家庭背景、教育背景等因素而表现出或大或小的差异。在涉及语句合法性(grammaticality)和适当性(appropriateness)判断的时候,自拟的例子往往是一些结构复杂、语义晦涩、日常生活中极少碰到的语句,而生成学派的创始人 Chomsky 与其同道偏偏十分喜好这样的例句。他们认为,正是这些常人看来复杂得不可思议的语句,最能反映说话人的语言能力,反映语言的本质特征。研究者对这类句子的判断是否可靠,是否会受个人语言(idiolect)的影响而缺乏足够的代表性,一直有很大争议(参见 Schütze 1996/2016)。Gibson & Fedorenko (2013) 认为,在句法和语义研究中,语句的合法性判断不能仅仅根据单个语言学家的语感,而是应该采用科学实验的常规方法,请许多人对有关语句是否能说作出独立的判断。他们提出,参与测试的人员数量要足够大,而且不能是语言学家,避免因专业背景的影响而作出不够客观的判断。Sprouse & Almeida (2013) 对此作出回应。他们采用 Gibson & Fedorenko (2013) 建议的方法,做了几个实验。其中一个实验是以 Adger (2003) 一书中的例句作为检验对象。该书共有 469 个例句,用以说明 9 个领域中 365 个句法问题。作者请了大量非语言学家对这些句子的合法性给出独立判断,结果显示,365 个句法问题中,参加测试的人员在 359 个问题的判断上同该书作者 Adger 一样,重合率达 98%。另一个实验是,生成语言学主要刊物 *Linguistic Inquiry* 2001—2010 十年间刊登的所有论文中,一共出现 1743 个例句,作者随机抽取了 292 个例句,涉及 146 个语言现象,实验请来的测试人员在 139 个现象的判断上同原作者一致,重合率达 95%。Sprouse & Almeida (2013) 根据这些测试结果得出的结论是,应该认为,句法和语义研究中,研究者根据内省获取的证据是可靠的。两位作者同时指出,有些例句介乎合法与不合法之间,文献中一般在句子前用问号标注,多人参与的检测结果能就这类句子的合法性给出更为全面的定量分析,而单个研究者则无法提供这类信息。

二、现代汉语语法研究代表作所用例证的来源

从现代语言学的角度研究汉语语法,始于 1898 出版的《马氏文通》。20 世纪前半叶的汉语语法研究,可以商务印书馆 1980 年代重印的《汉语语法丛书》十种,再加上胡以鲁和刘复的两部著作作为主要代表:



- 《马氏文通》 马建忠 1898
《高等国文法》 杨树达 1920
《国文法草创》 陈承泽 1922
《国文法之研究》 金兆梓 1922
《新著国语文法》 黎锦熙 1924
《中国文法论》 何容 1942
《中国文法要略》 吕叔湘 1942—1944
《中国文法革新论丛》 陈望道等 1943
《中国现代语法》 王力 1943
《汉语语法论》 高名凯 1948
《国语学草创》 胡以鲁 1913
《中国文法通论》 刘复 1920

以上十二部著作中,大多数研究对象是古代汉语。比较系统、全面地描写汉语语法的专著共有五种,分别是《马氏文通》《高等国文法》《新著国语文法》《中国文法要略》和《中国现代语法》。《马氏文通》和《高等国文法》研究对象是古代汉语,两部书有共同的特点:一是均从清儒的训诂学著作,如《助字辨略》《经传释词》《古书疑义举例》等中采撷了许多观点与材料;二是都因依西文语法著作的理论框架,参伍比较,为汉语语法搭建新的描写架构。在大量使用源自西文语法著作的概念、术语和分类组织框架的同时,两位作者也致力于揭示汉语本身的特点。用杨树达的话来说,“酌采欧西文法之规律,而要以保存国文本来面貌为期”(杨树达 1920/1955:2)。几乎所有例子,都标明出处。两书在解释词语意义时,很少提到现代汉语。《马氏文通》只有几处将古汉语词语同所谓“俗语”相比,如讲到“矣”字时,解释说“‘矣’字者,所以决事理已然之口气也。已然之口气,俗间所谓‘了’字也。凡‘矣’字之助句读也,皆可以‘了’字解之。”(吕叔湘、王海棻 1986:566)同书中所收七千多古汉语例句相比,提到现代口语的次数极少。

黎锦熙的《新著国语文法》是第一部以现代汉语为研究对象的语法著作。《马氏文通》和《高等国文法》受拉丁文语法著作的影响,均以词法为主,而《新著国语文法》主要以英语语法著作如 Reed & Kellogg 的《高级英文法》(*Higher Lessons in English*)为参照对象,句法占了重要地位,并提出后世影响较大的“句本位”主张。作者说,该书的研究对象是现代的汉语,不是文言文,是大众的普通话,不是某一阶层的行话和某一地区的方言。作者在解释书名中的“文法”一词时特意说明,文法就是 grammar 的翻译,不作“行文之法”讲,因为说话也要符合文法。之所以不用“国语法”,是因为当时小学有“说话”课,教“话法”,又称“口语法”,同“读法”(读书)和“缀法”(作文)对立,如果用“语法”或“白语文法”,可能会引起误解。黎锦熙同时说明,本书的主要研究对象,并不是口语,而是局限于用汉字书面化的白话文。当时流行的白话文中,欧化文体在知识阶层中相当通行,尤其常见于翻译作品和一些文艺作品。《新著国语文法》对书面语中的欧化现象举例很少,因此未能全面反映当时白话文语言特点的全貌,作者后来认为这是一个缺点。书中所收例句,主要出自明清传统白话小说如《水浒传》《西游记》《儒林外史》《石头记》《老残游记》等,以及五四以来新出的白话文作品,前者给了出处,后者则一般从略。

吕叔湘 1942—1944 年间出版了三卷本《中国文法要略》,上卷讲词语结构,中卷和下卷从



范畴和关系出发,研究它们的表现形式。上卷从形式出发研究意义,中/下卷从意义出发,研究表现形式,是第一部全面研究汉语语义的著作。该书的另一个特点是将文言和白话放在一起讲,读者通过对比,可以在文言和白话的异同以及联系方面获得深入了解。书中例句来源很广,大都取自当时流行的中学课本,从四书五经、传统白话小说、诗歌、词曲,一直到当代作家如冰心、老舍、徐志摩等人的现代白话文作品,大都标明出处。

王力 1943 年出版《现代汉语语法》。该书以不自造例句为原则,同时为避免语言夹杂,南腔北调,例证尽量先从用北京话写成的《红楼梦》里找,而且挑选口语化的句子,如家常谈话时所用,文绉绉的语句不用。如果《红楼梦》里找不到合适的例子,则临时自造,或从《儿女英雄传》里寻找。凡是书证,都标明出处。

就语料的性质来说,成书于 20 世纪上半叶的这三部研究现代汉语的语法著作有两个共同的特点,一是它们语料的主要来源都是白话文著作,包括传统白话文和现代白话文,是书面语而不是口语,大都标明出处;二是自拟的例子很少。

1948 年哈佛大学出版社出版赵元任的 *Mandarin Primer: An Intensive Course in Spoken Chinese* (《国语入门》)。这是赵元任在哈佛大学教美国陆军汉语口语短训班用的教科书。李荣将该书部分内容编译成一本书,1952 年以《北京口语语法》为名由开明书局出版。全书例句均为当代北京口语,没有标明出处。

赵元任又于 1968 年出版 *A Grammar of Spoken Chinese* (《中国话的文法》)。他在该书前言中写了这样一段话,解释为什么不用“国语文法”这样的书名:

“国语有标准官话的意思,而我想强化的是淡化官话的标准性,因为尤其是就语法而言,本书有关官话的讨论大都适用于整个汉语,甚至适用于文言的许多方面。”

(“*Gwoyeu* connotes Standard Mandarin, and my emphasis has been to de-emphasize the standard aspect of Mandarin, since, especially in matters of grammar, most of what is said here about Mandarin is true of all Chinese, even of a good part of the literary language.”)

(Chao 1968: viii)

赵元任在自己两本著作的中文书名中没有用“口语”或“汉语语法”这些词,认为它们是“新造的学术词汇”,有“专门名词的气味”。赵元任(1968)有吕叔湘的节译本《汉语口语语法》(1979 年由商务印书馆出版),香港中文大学出版社 1980 年出版丁邦新的全译本《中国话的文法》。书中例句有多种来源,主要是自拟的例子、真实口语记录以及口语化的当代文学作品,一般不标明出处。

1961 年,商务印书馆出版丁声树等撰写的《现代汉语语法讲话》。该书根据 1952 年 7 月至 1953 年 11 月《中国语文》连载的《语法讲话》修订而成。该书前言中说,书中“讨论的现代汉语,口语和书面语并重,口语以北京话为主,书面语以现代作品为主”。所用例句百分之八十以上采自当代白话文作品,并给出了来源,代表以北京话为主的口语例句不到百分之二十。朱德熙 1982 年出版的《语法讲义》,标明书面语例句出处的不到百分之五。1984 年商务印书馆出版吕叔湘主编的《现代汉语八百词》,该书编纂过程中参考了大量从书面材料里摘抄下来的例句,但在书中都没有标明出处。1990 年代以后,《中国语文》等学术刊物刊登的现代汉语语法论文,例句自拟的比例相当高,摘自书面材料的例句许多也并不标明出处。

如无其他限定词,“现代汉语语法研究”中的“现代汉语”,指的是现代汉语标准语。从上面的简单回顾中可以看出,从黎锦熙(1924)开始,1950 年之前的现代汉语语法研究所用例

证几乎全部出自用汉字书写的白话文,包括早期白话文和现代白话文作品,而且大都标明出处。自拟的例子和日常生活中的口语一般来说不足为证,因此在书中出现次数很少。这种做法一直延续到1950年代。1955年公布的普通话定义有两个部分:以北京语音为标准音,以北方方言为基础方言;1956年补充了第三部分,以典范的现代白话文著作作为语法规范。其中第一部分讲语音,第二部分讲词汇,第三部分讲语法。普通话的定义反映了当时的共识,即现代汉语语法研究所需的证据应该从书面语而不是口语中获取。

从1960年代开始,尤其是近三四十年来,通过自拟取得例证的做法渐渐为较多的研究者所采用,口语语料得到更多的重视,书证标明出处的做法也越来越松动。不过,多年来搜集例证最常用的还是传统方式,研究人员许多时间花在从各类书报杂志上搜集例句,打勾画线,转录为卡片,分门别类存档,写论文时再将相关卡片全部摊开。使用例句卡片,主要目的是便于研究,大量相关例句排列在一起,有助于对照、比较、联想和总结,从中找出语法规律。这样做的另一个好处是方便,可将例句直接写到论文里去。因此,从过去到现在,卡片一直是语法研究中普遍使用的工具。古代汉语和近代汉语研究人员写论文,例句必须标明出处;现代汉语语法研究论文所用例句虽然没有强制说一定得提供出处,但相当长一段时间里比较普遍的看法似乎是,标明出处比不标明出处显得更为严谨一些,哪怕是“我抱着花,向大家敬礼”这类几乎人人日常生活中都会说、毫无争议的句子。对普通例子都要标明出处,或许反映自拟的例证在研究者群体中可能公信力不足,要有书为证才好。这同时也反映,对于什么是现代汉语语法研究中的合格证据,研究者群体还缺乏足够的共识。

在现代汉语语法研究的证据问题上,我们在相当长的一段时期内主要依靠书面语而不是口语材料,较少使用自拟的例句。这种态度和做法的原因是什么,结果是什么,是我们下面要讨论的问题。

三、现代汉语标准语、口语、书面语

我们称之为普通话的现代汉语标准语,1950年代以前叫国语,如赵元任(1948)中文书名《国语入门》所示。明确定位为现代汉语标准语的国语,其实历史并不长。受日本明治维新后语言文字改革取得很大成绩的启发,从19世纪下半叶开始,汉字、汉文及汉语必须改革以促进中国社会的进步,成了当时社会精英阶层的共识。除了文字改革以外,清末开始的国语运动有两个主要目标,首先是言文一致,用接近口语的白话文取代传统文言文,成为主要的书面语文体;用来取代文言的新式白话文,到底应该是什么样子,同传统白话文有何区别,同日常口语有何联系?在这些问题上,学界讨论十分热烈。傅斯年(1919)在谈到怎样做白话文时,提出两个胡适认为最重要的主张,一是必须根据我们说的活语言,先讲究说话,话说好了,自然能做好白话文,二是要欧化,充分吸收西洋语言的细密结构,这样才能传达复杂的思想,以因应新时代的新需要。第一个主张听上去很有道理,但在当时实行起来会有很大困难。对于生活在广大南方方言区的人来说,除非他们先学会说比较流利的北方官话,否则无法将他们的活语言笔之于文。提出“我手写我口”的黄遵宪,母语是广东话,“我手写我口”对他来说,实际上是很难做到的。

国语运动的另一个主要目标是国语统一,向全国推广标准口语。中国现代教育制度二十世纪初才开始渐渐在全国建立。当时中国存在极大的方言歧异,缺乏普及的共同语,给不同地区的口语交流造成很大的障碍。作为标准语的国语应该是什么样子,也是一个引起热烈讨论

的问题。代表当时语言文字改革中坚力量的国语运动积极分子,于1916年成立了自己的组织“国语研究会”,1917年在《新青年》杂志上刊登了下面的主张:

同一领土之语言皆国语也。然有无数之国语较之统一之国语,孰便?必曰统一为便;鄙俗不堪书写之语言,较之明白近文,字字可写之语言,孰便?则必曰近文可写者为便。然则,语言之必须统一,统一之必须近文,断然无疑矣。

使立定国语之名义,刊行国语之书籍,设一轨道而导之,自然渐趋于统一,不过迟速之别而已。

不必虑统一之难,当先虑统一之无其术与具尔。同人等有见于此。思欲达统一国语之目的,先从创造统一之术与夫统一之器具为入手方法(严修等1917)

“国学研究会”认为,口语含有许多鄙俗成分,因此不能将口语等同于国语。哪些口语成分能成为“统一之国语”的一部分呢?条件是必须要“近文”,要“字字可写”,也就是说,能否以文字形式进入书面语,是口语成分是否有资格进入国语的衡量标准。

胡适(1918)提出“国语的文学,文学的国语”,简洁、明确地表述了他在现代汉语标准口语和标准书面语两者之间关系问题上的主张。胡适认为,“文学”,即书面语,要基于标准国语,用汉字将国语写成书面语,就是现代白话文,因此他将“白话文学”正名为“国语文学”。而另一方面,“国语”又要基于书面语,是现代白话文的口语形式。仔细分析胡适的主张,可以发现,这实际上是个先有鸡还是先有蛋的问题。没人能够根据他的主张,写出基于全国统一的标准国语的白话文,因为当时国语还没有标准;而要有标准国语,又先要有普遍认可、普遍使用的现代白话文。胡适说,不能等有了标准国语以后才开始写白话文,大家应该先做起来,慢慢形成标准。据他自述,他的白话文是从《儒林外史》《红楼梦》《儿女英雄传》《老残游记》等传统白话小说中学来的,同当时的国语没有多少关系。因此,胡适做的白话文,与其说是“国语的文学”,不如说是“《儒》《红》《儿》《老》的文学”(详细讨论参见陈平2017b)。取代文言文成为通用文体的现代白话文应该比较接近口语,通俗易懂,对此大家没有异议。但是,自1917年白话文运动起,直到现在,对于什么是现代白话文,我们实际上并没有普遍认可的标准。在如何对待文言、俗语、方言、明清小说中的传统白话、欧化文体等问题上,作家们基本上是各行其是,只要有读者就行。有过几次白话文标准的大讨论,涉及如何看待现代白话文同文言、方言、欧化文体和大众口语的关系,但离取得共识还差得很远(陈望道等1943)。现代汉语书面语在处理这些语言成分的实际做法上,同其他语言如英语很不一样。我们阅读今天的英文书报,见不到古英语的踪迹,但古汉语词语句式却是现代汉语书面语的常用成分。

确定国语的标准音,相对来说比较简单,1932年采用新国音的《国音常用字汇》公布以后,以北京音为国语标准音的原则基本上就此确定了。不过,即使有标准国语,要推广普及到全国也并非易事。自1950年代起完整的教育体系在全国建立起来以后,我们以举国之力,花了近半个世纪的时间,才将普通话普及到近百分之六十的人口中去。对于许多人来说,掌握国语和现代白话文的顺序,往往不是先学会说国语,然后根据国语做白话文;而是如胡适那样,先从书面语学会写大家能懂的白话文,不必理会是否符合标准国语。

国语长期缺乏普遍认可的标准并且未能在全国及时普及,造成学界对于什么是现代汉语标准语长期缺少共识,语法学家对于包括自己在内的语感缺乏信任,加上严重的方言歧异,知识阶层传统的重文字、轻口语——所有这些因素加在一起是1920年代开始的大半个世纪里,现代汉语语法研究倚重现成的书面语材料、较少使用自拟例证的主要原因。

主要以典范的现代白话文作品中出现的例子为经验证据,由此寻绎现代汉语标准语的语法规律,这既是我们定下的原则,也是自黎锦熙(1924)之后半个多世纪里的常规做法,这样的指导原则和实践给现代汉语语法研究带来什么问题呢?

首先,何为“典范”,何为“现代白话文”?两者都是模糊概念,伸缩范围可以很大,方言、古汉语、早期白话、欧化等语言成分,都会在现代汉语白话文作品中程度不同地出现,语法研究者对这些成分如何辨识和处理,取决于研究者本人的语言能力、理论水平和研究目的。涉及一个个具体例证时,研究者本人的语感是否可靠,基于语感所作的判断是否具有足够的代表性,能否为汉语语法研究者群体所普遍接受?我们常说“有书为证”,其实,仅仅是以文字形式出现过,这本身并不足以为证。根据来源多样、成分复杂的语料得出来的语法规则,其科学价值如何?这些都是我们常常思考的问题。根据大型计算机语料库进行现代汉语语法研究,因为语料库所收材料往往相当芜杂,情况可能会更为复杂。很难想象如“了”“着”“过”的用法等这样需要均质性很强的语料才能做出精准判断的研究课题,能够完全依靠大型计算机语料库提供的证据来完成(陈平 2015a、2015b)。现代汉语语法研究论文因作者对文中例句的可接受度判断与他人不同,从而引起争议,是时时发生在我们周围的事情。前面提到的 Sprouse & Almeida (2013) 那样的测试,我们还没有人做过。是否现代汉语语法研究领域里也会有那么高的重合率,是一个值得考察的问题。

朱德熙(1987)主张,要使现代汉语语法研究深入下去,应该对口语语法和书面语法分别进行细致的研究,观察它们之间的异同。以书面语为对象的语法著作应当将目前书面语语法的不稳定性恰当地反映出来。他同时提出,现代汉语语法研究应该以北京口语语法研究为基础。他用来作为范例的是赵元任(1968)《中国话的文法》。该书以北京口语为主要研究对象,从北京口语语料中寻绎汉语语法的规律。朱德熙(1987)提出的无疑都是很有见地并且十分及时的主张,其基本出发点是作为语法研究的语料,数量要足够大,纯度应该尽量高,即内部均匀、无矛盾。

北京口语的范围应该比典范的现代白话文著作窄得多,语法研究因语料内部不一致而造成的问题也会相应少得多。但是,实际研究过程中还是有不少有待回答的问题,有些是技术问题,有些是理论问题。首先,怎样搜集例证?可以通过自拟,也可以依靠北京人日常说话的录音或精确的录音转写。但是,许多因素决定了几乎不可能完全依靠这类录音材料开展比较大的研究课题。一般来说,日常口语有许多迟疑、省略、重复、倒错现象,如果做话语分析,这些现象应该都是研究对象(陈平 2017a),但要从中总结出语法规律,则需研究者决定哪些现象能说明语法规律因而可用,哪些不可用,这就取决于研究者本人的语感和理论水平,于是又回到是自拟可靠还是客观语料可靠的老问题。还有一个实际问题是,我们很难有大量可靠的录音和转写材料可用。几乎所有现代汉语语法研究者,包括赵元任和朱德熙本人,除了书证以外,涉及口语时实际上都主要通过自拟获取例证,必要时再采用调查或实验方式加以补充。从事现代汉语语法研究的非京籍语言学家应该都能说流利的普通话,但很少有人会说地道的北京话。如果承认非京籍语言学家也独立做出了大量的一流研究成果,能否就此推论,就语法规则来说,普通话和北京口语两者的交集非常大,对称差非常小,对于语法研究往往不产生实质性影响?不过,无论这个推论是否成立,许多从事现代汉语语法研究的非京籍研究人员,包括笔者在内,为审慎起见,研究过程中经常同时用上自拟、书证和调查三种搜集证据的方式,重要例句或者自己没有太大把握的地方,都要向北京本地人核实一遍才放心。

上面的讨论引出另外一个问题:京籍语言学家在现代汉语语法研究上无疑具有独特的优势,但正如朱德熙(1987)所说,汉语语法研究从来不是京籍语言学家的专利。那么,非京籍汉语语法学家,他们从事以北京口语为基础的语法研究时,语感从何而来?他们的普通话是怎么学的?中小学教育、广播影视应该是主要渠道,但另一个重要渠道是书面语。非北方官话区的人从书面语而不是口语中学会许多普通话常用的词语结构,一直是十分普遍的现象。甚至北京人也从书面语学会一些北京话原来没有的成分,渐渐用在日常口语中。黎锦熙1951年对于欧化长句,发表了这样的意见:虽然出现次数少,不适宜于口语,但如果符合汉语“共同的法式和基本的规律”,因为许多人实际上是通过书面语学习语言的,所以这些欧化长句尽管不是“从群众中来”,但很容易“向群众中去”,变成大家的普通说法(黎锦熙1924/1992:24-25)。这个判断不仅适用于欧化成分,也适用于方言成分。“有没有VP”是吴语成分,“有VP”是闽南方言成分,这些用法开始都是先在书面语中出现,没过多少年,“有没有VP”已经完全进入普通话,“有VP”也越来越常用(王森等2006)。这说明,书面语不仅是人们掌握北京话和普通话的重要渠道,也同时成为北京口语、普通话、方言、白话文等语体之间相互借用、相互影响的重要渠道,这就无可避免地降低了北京口语的独特性和“纯度”。

造成北京口语内部非均质性的原因,除了方言和欧化等外源成分的影响之外,还有两个重要因素,一是谁的北京口语,二是口语所指对象到底应该是什么。

北京人本身是个相当模糊的概念。年龄、职业、居住地区、家庭和教育背景不同的人,其日常口语也不会完全相同。老舍和王朔是现代汉语语法研究中最常用来作为北京口语发话人代表的两位作家,他们两人的语言还是有相当差别的。某些数词如“一”“二”等后跟名词时省略个体量词,是当代北京口语中的常见现象,青年人中尤其盛行。这种用法据说是王朔开的头,他的小说和根据小说改编的影视作品走红以后才流行开来的。同样代表北京口语但年代早一些的作品,如老舍的《茶馆》,没有或极少见到这样的用法(陈平2016)。王朔是北京军队大院子弟,从小在封闭环境中长大,说的是大院通行的普通话,同北京老百姓几乎没有什么接触。据王朔自述,上中学开始是在东城区,后来转到西城一所中学后才第一次听周围的人说北京话,开始许多词根本就听不懂。我们的问题是,老舍、王朔、在京机关学校工作人员、内九城胡同里的居民、京郊的工人农民,谁的语言能代表北京口语?还是他们都能代表北京口语?如果都能代表北京口语,那么北京口语和普通话到底有什么区别?北京人这个概念不易精准定义,他们的口语也同样如此。

根据使用场合、目的、受众等因素的不同,口语的用词、句式和组织也会有很大的区别。陶红印(1999)提出,就现代汉语语法研究的对象来说,口语书面语两分不够,应该根据语体(genre)进行更详细的分类,找出语体之间语法特征的异同。文中证明,现代汉语的一些语法现象,如“把”和“将”的用法,用语体的概念比用口语书面语的区别更能将有关规律揭示出来。这告诉我们,有必要更深入地探讨“北京口语”或“口语”“书面语”这些概念的内涵和外延,以及它们同语体等概念之间的关系。从例证采集方法上来看,涉及语体时,只能更多地利用自拟之外的其他两种方法,主要还得依靠书面语材料。

朱德熙(1987)文中提出的另一个建议是对口语语法和书面语语法分别作细致的研究,找出它们的异同。这也是一个很好的建议。限于篇幅,我们另文讨论对该建议的具体想法。与此同时,我们不妨观察一下英语语法研究中的相关情况。近几十年来最重要的英语语法著作应该是Huddleston & Pullum(2002)。它是一部1842页的巨著,出版后广受好评。作者说,他

们要写一部尽量在口语和书面语之间保持中立的著作,但读者可能还是会觉得,该书在语料问题上偏重书面语。作者解释说,如果确实对书面语有所偏重,那并非他们的编写原则,而只是出于实际考虑:书面语能直接引用,书面语语料比口语语料容易获取,口语材料如录音中错误很多等。更重要的是,作者持有这样一个假设(assumption):口语和书面语各自内部都有文体(style)上的差异,但口语句法和书面语句法之间的显著分歧罕见到不存在的程度(rare to the point of non-existence)(Huddleston & Pullum 2002: 11-13)。也就是说,根据书面语提供的例证,写一部在口语和书面语之间保持中立的英语语法著作,理论上是做得到的。我认为,英语语法学家能将撰写大部头的语法著作建立在这样的假设之上,主要原因之一是现代英语(Modern English)历史较长,一般认为从1776年开始就基本定形了。我们现代汉语语法学家能持这样的假设吗?

四、结 语

本文结束之前,我引用一段黎锦熙1951年讲的话(黎锦熙1924/1992: 21):

在三十多年以前,一般人只注意中国的文言文法,对于口语和白话文,听说也要建立起文法来,就感觉到是一件新奇而困难的事:你究竟根据一些什么材料呢?

一个世纪过去了,我们今天依然问同样的问题。从现代汉语语法研究实践来看,1950年代之前的现代汉语语法著作,根据的几乎全部是书面语材料,包括早期白话文和现代白话文作品,并且大都标明出处,以示有书为证。这种做法延续到1950年代。从1955—1956年公布的普通话定义也可以看出,它代表了当时的共识。1960年代之后,尤其是近三四十年来,例证自拟的做法渐渐通行开来,更为重视口语,尤其是北京口语的研究,从书面语直接引用或改写的例句也可以不必注明出处。这些变化反映汉语语法研究者对自己语感的自信在不断增强,反映现代汉语语法学界对于什么是合语法或不合语法的用法有了越来越多的共识,也反映现代汉语标准化程度在不断提高。

现代汉语语法研究应该如何获取证据,涉及许多同现代汉语标准语的起源与发展有关的问题,主要围绕如何看待作为语料的现代汉语标准语、口语和书面语这三者之间的关系,如何把握自拟、利用真实语料以及实验/调查这三种获取证据的方式各自的特点。本文的主要目的是梳理有关问题的来龙去脉,并指出某些问题的症结所在。本文同时提出许多问题,没有现成答案,希望引起更多的关注。

参考文献

- 陈 平 2015a 系统中的对立——谈现代语言学的理论基础,《当代修辞学》第2期。
陈 平 2015b 从现代语言学经典论著看语言学论文的写作与发表,《当代修辞学》第6期。
陈 平 2016 汉语定指范畴和语法化问题,《当代修辞学》第4期。
陈 平 2017a 话语的结构与意义及话语分析的应用,《当代修辞学》第2期。
陈 平 2017b 白话文运动:传承还是摒弃,《语言战略研究》第4期。
陈望道等 1943/1987 《中国文法革新论丛》,商务印书馆。
丁声树等 1961 《现代汉语语法讲话》,商务印书馆。
傅斯年 1919/1935 怎样做白话文 《建设理论集》,胡适编选,上海良友图书印刷公司。
胡 适 1918/1935 建设的文学革命论 《建设理论集》,胡适编选,上海良友图书印刷公司。



- 黎锦熙 1924/1992 《新著国语文法》,商务印书馆。
- 吕叔湘 王海葵 1986 《马氏文通读本》,商务印书馆。
- 陶红印 1999 试论语体分类的语法学意义,《当代语言学》第3期。
- 王森、王毅、姜丽 2006 “有没有/有/没有 + VP”句,《中国语文》第1期。
- 严 修等 1917 中华民国国语研究会暂定简章,《新青年》第3卷第1号。
- 杨树达 1920/1955 《高等国文法》,商务印书馆。
- 朱德熙 1987 现代汉语语法研究的对象是什么?《中国语文》第5期。
- Adger, David 2003 *Core Syntax: A Minimalist Approach*. Oxford University Press.
- Chao, Yuen Ren 1948 *Mandarin Primer: An Intensive Course in Spoken Chinese*(《国语入门》). Harvard University Press.
- Chao, Yuen Ren 1968 *A Grammar of Spoken Chinese*(《中国话的文法》). University of California Press.
- Gibson, Edward, & Evelina Fedorenko 2013 The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28.1/2: 88-124.
- Huddleston, Rodney, & Geoffrey K. Pullum 2002 *The Cambridge Grammar of the English Language*. Cambridge University Press.
- Schütze, Carson T. 1996/2016 *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Language Science Press.
- Sprouse, Jon, & Diogo Almeida 2013 The empirical base of data in syntax: A reply to Gibson and Fedorenko. *Language and Cognitive Processes*, 28.3: 222-228.

On Empirical Evidence in the Studies of Modern Chinese Grammar

Chen Ping

School of Languages and Cultures, the University of Queensland

Abstract: Empirical evidence used in studies of Modern Chinese grammar is derived from three major sources: self-made examples, existing raw or primary data of spoken and written Chinese, and data collected via fieldwork or experimentation. Since *The First Comprehensive Grammar of Modern Chinese* written by Li Jinxi in 1924 until the 1960s, the overwhelming majority of illustrative examples used in studies of Modern Chinese grammar were drawn from written materials with their origins explicitly indicated, only to be occasionally complemented by self-made sentences. The principle of evidence collection is also confirmed in the definition of Putonghua promulgated in 1955-1956 which states that the standard form of Modern Chinese looks to exemplary works in *baihuawen* (vernacular literary language) for its grammatical norms. It is only since the 1980s that self-made examples, and examples from spoken vernacular, have gradually increased in writings on Modern Chinese grammar. Lack of agreement on grammaticality judgements on illustrative sentences has also prompted the suggestion that studies of Modern Chinese grammar should be based primarily on the Beijing vernacular, which provides a more homogeneous data base than the vernacular literary Chinese. This article undertakes an analysis of major issues and controversies surrounding collection of empirical evidence in studies of Modern Chinese grammar. It explicates how the way in which modern spoken and written Chinese has been developed and promoted since the early 20th century has been a determining factor underlying the selection of illustrative examples in writings on Modern Chinese grammar.

Keywords: Modern Chinese grammar, Putonghua, Beijing vernacular, spoken Chinese, written Chinese, Vernacular Literary Language

