

对外汉语教材词类标注问题研究 ——以《发展汉语》(第二版)综合教材为例

杨旭 付敏

(复旦大学中文系,中国上海 200433;四川外国语大学中文系,中国重庆 400031)

摘要: 文章结合国内外学界在词类和词类标注研究方面的最新成果,考察了《发展汉语》(第二版)综合教材的词类标注,发现存在漏标、错标和多标三类错误;此外还存在与词类标注相关的释义与词类之间形类不一致的问题。这些问题的产生多是由于不正确地参考语文词典造成的。我们既要认识到对外汉语教材和语文词典从根本上隶属于不同的层面,也要认识到两者存在双向互动,尤其是对外汉语教材中存在类似于词典的功能块——生词总表,所以要在承认“唯一性”为首要原则的基础上,做到生词表自始至终遵循“唯一性”原则,而生词总表则遵循“两步走”原则。

关键词: 《发展汉语》(第二版);词类;词性标注;对外汉语教材;兼类

中图分类号: H195.4

文献标志码: A

文章编号: 2221-9056(2018)05-0072-07

DOI 编码: 10.14095/j.cnki.oce.2018.05.008

一、引言

在现代汉语词类范畴化大讨论中,对外汉语教材(尤其是兼类词)的词类标注^[1]备受关注,是学界讨论的难点和热点问题(方清明 2007;马宏基 2008;王弘宇 2001、2013;龙煌汕 2013)。对外汉语教材是否有必要标注词类的问题已经没什么异议了。根据 Schmidt(1990)提出的关注假设,二语习得需要更多的显性语法知识,其中标注词类无疑有利于留学生正确理解词的意义、掌握词的用法。我们在教学实践中也发现,词类偏误是留学生最基本的偏误类型之一。比如“我从小就兴趣踢足球”把名词“兴趣”等同为动词“爱好”,“老师搞教育了我”则不知“教育”兼属名词和动词,只知道其名词属性。教材只有提供词类知识才可能帮助学生克服这类偏误。尽管如此,针对指导原则(“唯一性”“全面性”还是“两步走”)仍有一些争议,尤其是体现在没有清醒地认识到对外汉语教材和语文词典在词类标注上的关系。有鉴于此,我们拟结合国内外学界在词类及词类标注研究方面的最新成果,考察《发展汉语》(第二版)综合教材(包括分初、中、高三个层级各2册,后简称

收稿日期: 2017-07-11

作者简介: 杨旭,复旦大学中文系博士生,研究方向为现代汉语语法。Email: yangxu1216@hotmail.com

付敏,四川外国语大学研究生,研究方向为对外汉语教学。Email: 752217683@qq.com

基金项目: 本研究得到了重庆市研究生科研创新项目(CYS15179)的支持。《海外华文教育》匿名审稿专家和王仁强教授就本文提出过宝贵意见,项目小组成员牛瑶和吴冰霜参与了《发展》词类标注的考察工作,特此致谢。文中不妥之处概由本人负责。

“《发展》”) 在词类标注实践中存在的问题,在此基础上厘清对外汉语教材和语文词典在词类标注上的关系,以求为对外汉语教材编者带来启示,从而进一步提高对外汉语教材的词类标注水平。

二、《发展》词类标注问题考察结果

在《发展》的“基本使用信息”中,高级综合的使用水平为“已掌握 3500-4000 词汇量”。我们进行了统计,发现共有 4291 个生词。这些生词的词类标注情况如表 1,可见在这 4291 个生词里,有 3931 个生词(占比 91.61%) 被识别为词,并被标注了词类。通过逐一考察,发现存在三类问题:漏标、错标和多标。此外还存在一类相关问题,即释义与词类之间的形类不一致。

表 1 全部生词的词类标注情况

类别	未标注词类		标注词类	合计
	非词语素	短语		
数量	5	355	3931	4291
比例	0.12%	8.27%	91.61%	100%

(一) 漏标

我们根据《现代汉语词典》(第 7 版)(简称《现汉》)考察了未标注词类的生词,发现有 29 个生词应标注词类而没有标注,包括“搬家、不用、上网、刷卡、算了、不得了、金字塔、随着、出门、看病、出去、出院、对不起、结账、跑步、起床、散步、上班、生病、住院、打的、想不到、学好、拜年、串门儿、抬头、不像话、起劲、赘言”,如表 2(因篇幅限制只列了 10 个)。

表 2 《发展》漏标而《现汉》标注的生词(部分)^[2]

生词	语料	《现汉》词类标注
搬家	我已经搬家了/山田搬家了	动词
不用	山田:谢谢你!王朋:不用谢	副词
上网	晚上我学习汉语,有时候看电视、上网	动词
刷卡	在中国,用银行卡很方便,可以存钱、取钱,买东西、住宾馆也可以刷卡	动词
算了	算了,咱们别找了	助词
不得了	黛比第一次吃油条,喜欢得不得了	形容词
金字塔	我的面前就是金字塔	名词
随着	随着“文革”的结束	介词
出门	出门的时候多穿点儿衣服,平时多吃水果,多喝水	动词
看病	你去医院看病了吗	动词

漏标包括两种可能的原因,一是误将词识别为短语,如“搬家、出门、看病”等;二是回避了新词的词类标注,如“上网、刷卡”等。第一点情有可原,因为词和短语有时难以区分,我们只能通过基于语料库的使用模式调查来判断其词化程度。对于第二点,“总前言”里明确表示会积极吸收现代社会中已经非常普遍的新词(如“手机、打折、刷卡”等)以示教材的时代性,但是如何为吸收进来的新词标注词类却未作说明。我们认为回避词类标注不值得提倡,只会增加学生的疑惑,导致偏误的产生。

(二) 错标

首先是错将语素标为词。“啤花泉水成佳酿，酒自清醇味自芳”(高级综合(Ⅱ)第8课)中的“酿”被标为了名词。但是基于 CCL 现代汉语语料库的调查表明，“酿”只出现在“佳酿”“西洋名酿”和“芳酿”等少数结构中，其个例频率和类型频率都很低，说明意为“酒”的“酿”在现代汉语中不成词。有鉴于此，我们建议把“酿”处理为语素，至于其在古代汉语中的成词用法可以加注说明。再如把“大堂哥叫‘三午’”(高级综合(Ⅰ)第2课)中的“堂”标为了名词，明显不当。

其次是错标为他类。这类生词共有 35 个，包括“遍、事先、堂、数码、灰色、本能、同窗、造反、常规、萌芽、供奉、独身、沿海、丛、罐、得、感冒、难过、处罚、扫兴、患难、迟疑、黯然、羞辱、感慨、刻度、水文、类似、流行、伤、致辞、真心、变故、并、那么”。对这类错误简单分类，结果如表 3：名词被标错的最多(9 个)，动形副次之(6 个)；被标为名词、动词的词性类型最多，都是 5 种。名动形副之所以容易出现混淆，可能是受制于汉语词类具有多功能性的主张，即“汉语词类与句法成分之间不存在简单的一一对应关系”(朱德熙，1985: 4-5)。该观点认为名词可以做主宾语、谓语(一定条件下)和定语，导致本来为名词、形容词、副词的词被错标为名词。下面举例分析：

表 3 《发展》中错标为他类分类表

正确标注 错标	名	动	形	副	助动	量	介	构词
名		6	1	2		2		2
动	8		5	1	1			2
形	1			2				
副								
助动								
量								
连								1
代				1				
小计(35)	9	6	6	6	1	2	0	5

(1) 好像事先练习过。

“事先”被标为名词，可是释义时却用了副词性的“in advance; beforehand”。《现汉》把概括词“事先”标为名词，语委语料库中类似例句的标注则矛盾重重(被标为时间名词、处所名词、副词、普通名词的索引行都有)，但是基于 BCC 现代汉语语料库的使用模式调查发现，“事先+V/VP”是“事先”出现的主要结构(12326/19443=63.40%)，说明概括词“事先”的无标记功能是在动词或动词短语前做状语(即副词)，更不用说“好像事先练习过”中的个体词“事先”应被标为副词了。

(2) 曾听人说，这女人也一直独身。

“独身”被标为名词，释义为“不嫁或不娶的成年人”。这里的“独身”表陈述功能，意为“成年人没有结婚”，应为动词。这种用法较为普遍，基于 CCL 现代汉语语料库的使用模式调查发现，在 716 条有效索引中动词用法(“成年人没有结婚”)有 350 条，占比 48.88%，名词用法(“不嫁或不娶的成年人”)和副词用法(“只身”)分别只有 194 条和 72 条；数据证明“独身”的动词用法已经规约化。事实上，根本不需要大费周折地去调查，参考《现汉》就能避免这类错误，因为《现汉》已经收录了“独身”的动词用法。这证明了语文词典在对外汉语教材词类标注中的重要参考价值。

(3) 女孩儿很难过，他怎么也想不通。

(4) 又有些犹豫。

(5) 他红着脸，迟疑不动。

(6) 索重阳很扫兴,家里也没有热饭热菜等着自己。

以上四个词都被标为动词。它们出现在主语之后,语义表示心理状态,应是表陈述功能的陈述形容词(predicative adjective)。《现汉》也只收录了它们的形容词用法,不知《发展》参照了什么标准。此外,“扫兴”除形容词用法外,还可以出现在“扫了……的兴”这样的语境中,说明是一个离合词。在高级阶段的生词总表中,可把类似离合词处理为词和短语两个级阶,词需要标注词类,短语不需要标注词类。离合词为汉语词汇所特有,外国留学生的母语中普遍没有词语离合和合的现象。它本来就是对外汉语教学中的难点,如果能够对离合词进行“词”与“短语”的分级阶教学,并对其进行准确的词类标注,势必会减少学生负迁移现象的发生,达到更好的教学效果。

(7) 没有额外的、灰色的收入,这些人是不会安心的。

“灰色”被标注为名词,释义为“原指一种介于白与蓝之间的颜色,本文指来路不明的收入”。词性标注和释义均有问题。《现汉》把该义项标为“属性词”,释义为“比喻不明朗的;不正当的”,配例为“灰色收入”和“灰色市场”。建议标为形容词。《发展》的词类体系中没有属性词,可具体解释其属性词用法)。

(三) 多标

多标指的是教材中只出现了某词的一种用法,但却标注了该词的多个词类(即处理为兼类)。这类词共有20个,包括被标为名动兼类的“进步、解释、劳动、想象、比喻、顾虑、刹车、警卫、指令、骚乱、挑战、题刻、构思、决策、联想、酿、倡议”和被标为名形兼类的“黑暗、实惠”,以及被标为动形兼类的“麻痹”。各举几例,如表4。

表4 生词表中被标为兼类但课文中只出现了一类的生词

词性	生词	语料
名动	进步	李老师说,您觉得我们汉语有进步吗 别着急,你们请示多听、多说、多看、多写,进步一定更快 ^[3]
	想象	搭车并不像想象的那么容易
形名	黑暗	接完电话,我感觉广场一片黑暗
	实惠	尤其是那些已经得到了实惠、对金钱的力量有过体验的官员
动形	麻痹	他只有不断喝酒,让自己的心暂时麻痹,忘记那些孩子们

(四) 释义与词类形类不一致

双语词典词目翻译中有一条编纂原则叫“词类对译”,指尽可能使用和源语词词类相同的译语词(黄建华,1998:10)。该原则亦适用于单语词典和对外汉语教材的词目释义,可以称之为“形类一致”原则,即释义语言的形类要和词类标记相一致。《发展》初级综合用英文释义,中、高级综合用中文释义,在释义上大部分能做到“形类一致”,但也存在少数不一致的情况,如表5。

表5 不符合“形类一致”原则的生词

生词	词类	释义	释义形类
事先	名	In advance; beforehand	副
难过	动	be distressed	形
常规	名	普通的、一般的、通常的	形
水文	动	自然界中水的各种变化和运动	名
变故	形	意外发生的变化和事故;灾难	名
斗胆	形	大胆地,常用于谦辞、敬辞。斗是中国古代的一种量器。如:请允许我~说一句	副

三、对外汉语教材和语文词典在词类标注上的关系

对外汉语教材标注词类不可避免地要参考语文词典。《发展》在“总前言”中说“广泛参考了各类大纲和标准”,而各类大纲和标准都要参考语文词典,如《汉语水平等级标准与语法等级大纲》(1996)在“前言”列出的参考文献中就有四部语文词典,包括最权威的《现代汉语词典》(1979)。我们在上文提出修改建议时也参考了语文词典,可见对外汉语教材和语文词典的密切关系。但是参考语文词典容易出现一些错误,以《发展》为例,漏标是由于没有参考或采纳语文词典的词类标注,误标中有一些问题本来可以在参考语文词典后避免(如“独身”),而多标则是过度地参考了语文词典。尤其是多标的问题,是和词类标注原则最为相关的一个问题,也是过去讨论的最多的一个问题(王弘宇 2001; 方清明 2007; 吴甜田 2014)。如方清明(2007)所言,对外汉语教材编者以为在生词表中简单罗列生词的各种词性就可以大功告成了,这其实是受了语文词典标注词性的影响。尽管参考语文词典有一些弊端,但因此而把语文词典束之高阁就是因噎废食了。所以,语文词典和对外汉语教材在词类标注上到底是什么关系?恐怕只有厘清这个问题才能更合理地参考语文词典。

我们认为,两者既存在区别,也存在联系。首先,发生于对外汉语教材和语文词典中的词类范畴化隶属于不同的层面。王仁强指出,词类范畴化发生于两个层面,即个体词(token word)的词类范畴化和概括词(type word)的词类范畴化,前者是说话者的命题言语行为(包括指称、陈述、修饰等)的表达过程,后者体现为一个语言社群集体无意识的自组织过程,其核心是规约化/相变(Wang 2014; 王仁强、黄昌宁 2017)。与之对应,对外汉语教材是针对言语层面句法中的个体词进行词性标注,它只能隶属于一个词类范畴,不存在兼类词问题;而语文词典则是针对语言层面词库中的概括词进行词类标注,它可能属于一个或多个词类范畴,后者体现为兼类词。

其次,对外汉语教材和语文词典中的词类范畴化存在双向互动。一方面,概括词的词类属性是随着个体词在言语层面句法中(即各种命题言语行为构式中)反复使用而发生演变的,概括词的语法多功能/兼类与使用频率(含个例频率和类型频率)密切相关(Wang 2014; 王仁强、黄昌宁 2017),所以对外汉语教材中的词类范畴化可能会影响到语文词典中的词类标注。另一方面,因语文词典收录的都是规约化用法,凡教材中出现的词性多能在词典中找到,这使得语文词典成为对外汉语教材标注词性的重要参考文献。需要引起格外注意的是,在遇到言语层面句法中的词类活用现象时,需要编者通过个体词的句法功能与其意义事件或语境意义(meaning events)的关联标记模式来进行判断(Wang 2014; 王仁强、黄昌宁 2017)。例如“副了多年,想当正职”中的“副”,在语文词典中绝不可能找到其动词用法,但我们仍要实事求是地把它标为动词。

最后,两者也有小部分类似的地方。对外汉语教材中有类似于词典的功能块——生词表,但很多人忽视了有两类生词表,一类是一篇课文之后的生词表,一类是所有课文之后的生词总表。生词表提供与课文对应的词性标注及释义,满足学生一边阅读一边查询的需要,可谓是随文标注和释义;因只有一篇课文,故出现兼类的可能性较小。生词总表提供所有课文中的词类标注及释义,把所有生词汇总以方便学生查询与记忆;这时出现兼类的可能性较大,把它们所属的全部词类标注出来,有利于学习者系统地学习和掌握词汇。这也难怪有人把生词表喻为微型词典,其实针对的是“生词总表”。但要注意,词典是以概括词为中心,配例是用来佐证其用法的;而对外汉语教材是以课文(可谓是配例的集合)为中心,生词表扮演着比生词总表更重要的角色,切不可本末倒置。

明确了对外汉语教材和语文词典的关系,也就能够明确词类标注原则了。过去针对李晓琪(1997)提出的“准确性”有三派分歧观点,“唯一性”原则主张教材不同于词典全方位穷尽每个词的

词类,只能针对课文中出现的具体语境,不应该超出课文(王弘宇,2001;方清明,2007;吴甜田,2014)。“全面性”主张对外汉语教材应该标注生词的所有词性,这样有助于学生全方位地明白和把握词的意义和用法,只标注单个词性不利于学生建立完整的语法知识构架(马宏基,2008;李琳,2013)。“两步走”的折中方案主张初、中级阶段只标明生词在课文中的词类,高级阶段标注生词所有词类(胡佳敏、郝艳琴、潘婕龙,2012)。因对外汉语教材的词性标注属于言语层面,所以我们主张“唯一性”是首要原则,但因其中还存在类似于词典的生词总表,而且考虑到二语习得是一个循序渐进的过程,所以我们还主张在高级阶段的生词总表中遵循“全面性”原则。具体来说就是,一篇课文之后的生词表自始至终遵循“唯一性”原则;而对于“生词总表”,在初、中级阶段只标注课文中出现过的词类,在高级阶段则参照语文词典标注所有词类,即遵循“两步走原则”。

四、小 结

《发展》在词类标注上存在漏标、错标和多标三类错误,此外还存在与词类标注相关的释义与词类之间形类不一致的问题。这些问题多是由于不正确地参考语文词典造成的,具有混淆语言层面和言语层面、罔顾言语层面句法中的个体词词性而从语文词典照搬概括词词类的嫌疑。我们既要认识到两者从根本上隶属于不同的层面,也要认识到两者存在双向互动,尤其因为对外汉语教材中存在类似于词典的功能块——生词总表,所以要在承认“唯一性”为首要原则的基础上,做到生词表自始至终遵循“唯一性”原则,而生词总表则遵循“两步走”原则。《发展》中存在的问题源于现代汉语词类研究困境,本文对相关问题尤其是对外汉语教材和语文词典在标注词类上的关系做出了澄清,不仅有利于提高对外汉语教材的词类标注水平,或许也将为现代汉语词类研究带来启示。

注释:

- [1] 本文严格区分“词类”和“词性”两个术语:词类是概括词在语言层面词库中所隶属的范畴或类别;词性是个体词在言语层面句法中的语法性质或属性(可以参考杨旭,2017:25)。故词典中为“词类标注”,语料库中为“词性标注”,对外汉语教材的文内标注和文后生词表标注为“词性标注”,生词总表为“词类标注”,泛泛而言时统称为“词类标注”。
- [2] 因言语层面句法中存在词类活用的修辞现象,故个体词在语料中的词性不一定是概括词的词类。我们一一进行了核对,恰巧全部一致。
- [3] 虽然目前动词自指用法是否应标为名词存在争议,但这两条语料同为做主宾语,词类标注应该一致,或者名词或者动词。

参考文献:

- 方清明《对外汉语教材词性标注问题研究》,《暨南大学华文学院学报》2007年第2期。
- 胡佳敏、郝艳琴、潘婕《对外汉语教材中兼类词词性标注问题的调查研究》,《神州》2012年第11期。
- 黄建华《改进双语词典的翻译》,《辞书研究》,1998年第2期。
- 李琳《对外汉语教材词汇研究——基于两套对外汉语教材双音节动词的考察》,苏州大学硕士学位论文,2013年。
- 李晓琪《外国学生现代汉语常用词词典编纂散论》,《世界汉语教学》,1997年第3期。
- 龙煌汕《对外汉语教材中兼类词词性标注问题考察》,《北京地区对外汉语教学研究生论坛论文集》,2013年。
- 马宏基《对外汉语教材中兼类词的标注问题》,《教育实践研究》2008年第10期。
- 王弘宇《对外汉语教材词类标注的考察角度与处理原则》,《语言教学与研究》2013年第5期。

王弘宇《教材中词性标注的针对性原则》,《语言教学与研究》2001年第5期。

王仁强、黄昌宁《从双层词类范畴化理论看现代汉语自指词项的兼类问题》,《外国语文》2017年第1期。

吴甜田《对外汉语教材<汉语教程>(修订本)生词释义研究》,《语言应用研究》2014年第7期。

杨旭《〈现代汉语词典〉(第6版)兼类词表征策略研究》,四川外国语大学硕士学位论文2017年。

朱德熙《语法答问》,北京:商务印书馆,1985年。

Wang, R. Q. Two-level word class categorization in analytic languages: A comparative study of multiple class membership in Modern Chinese and Modern English. In Proceedings of the 36th Annual Conference of the German Linguistic Society, March 5th - 7th 2014, University of Marburg, Germany, 2014: 345-347.

A Study on Part-of-Speech Tagging in Chinese Textbooks for Foreign Learners

——Based on the *Comprehensive Course of Developing Chinese* (2nd ed.)

YANG Xu & FU Min

(Department of Chinese Language and Literature, Fudan University, Shanghai 200433 China;

Department of Chinese Language and Literature, Sichuan International Studies University, Chongqing 400031 China)

Abstract: This paper examines the part-of-speech tagging of the *Comprehensive Course of Developing Chinese* (2nd ed.) in view of the latest developments of word class and word class labeling or part-of-speech tagging study home and abroad, finding that there exists three problems, namely missing to tag, wrong tagging and redundant tagging. Besides, the inconsistencies between part-of-speech tagging and definition are pointed out. It is caused by the improper reference to Chinese dictionaries. We argue that Chinese textbooks for foreign learners and Chinese dictionaries belong to different levels but they also interacted with each other. Considering that total vocabulary list serves a dictionary-like function, we argue for the unicity principle in vocabulary list and the two-steps principle in total vocabulary list, admitting the unicity principle as the first principle.

Key words: *Developing Chinese* (2nd ed.); word class; part-of-speech tagging; Chinese textbooks for foreign learners; multiple class membership