

莱比锡标注系统及其在汉语语法研究中的应用

陈玉洁^{1,2} Hilario de Sousa² 王健²

倪星星² 李旭平² 陈伟蓉² Hilary Chappell²

(1, 浙江杭州 浙江大学 chenyujie0410@zju.edu.cn; 2, 法国高等社会科学院)

提要 语言行间标注(interlinear gloss)是进行语言学描写非常有用的工具,利用它记录的语料便于进行归类、储存和分析。但是这类标注方式在汉语研究领域还没有得到普遍应用。本文主要介绍目前对语言标注所普遍采用的莱比锡标注规则,并且结合法国高等社会科学院 Sinotype 项目组描写汉语方言语法时的处理方法,来说明莱比锡规则在标注汉语语法中的实际运用。

关键词 行间标注 莱比锡标注 汉语语法标注

壹 语言标注简介

语言标注在西方语言学历史上已经有相当长的历史,它原是因外语教学与研究的需要而出现的。标注系统又称为行间标注(interlinear gloss),指出现在原始文本(original text)和翻译文本之间的一系列对原始文本意义的描述和定义,简单地说,就是对原始文本进行逐字或逐词的说明。行间标注的主要功能是使读者能够追踪到原始文本和翻译文本之间的关系,并且可以非常明确地展示原始语言(也就是目标语言)的结构。

语言之间的直接互相翻译看不出不同语言间结构上和组合上的差异。比如下面两例对:

①锅铲——egg flip

②我昨天就把东西还给他了。——I gave it back to him yesterday.

从上两例可见,翻译实不足以看出需要翻译的语言中任意一个语言成分的功能、成分与成分之间的句法关系以及成分在句子中具体的句法位置。标注的目的就是显示这些被翻译所掩盖的语言成分之间的具体对应关系以及各个语言成分的意义,包括概念意义和语法功能。

行间标注长期以来服务于多种目的,比如在进行语言教育时用于对双语课本的注释,或者是语言学者利用一种元语言去描述一种陌生的语言。

行间标注系统一般包括下列几项内容,并按以下顺序出现:

- 1) 使用对象语言本身所使用的文字进行正确的拼写;
- 2) 把对象语言使用拉丁字母拼写出来,一般情况下,这项标注需要对象语言已经有固定的拉丁化拼写方案,比如汉语拼音系统。
- 3) 语音转写,使用国际音标标出语言的实际读音。
- 4) 形态音位学方面的标注,这里指的在语素上发生的语音改变,这一部分对于汉语来说特别重要。
- 5) 逐词或逐个语素进行标注。
- 6) 运用元语言进行自由翻译。

这6项组成了比较完整的标注系统,但是由于文章的空间限制,语言学论文或专注的标注往往都是有选择地选择其中一些项目或对这些项目进行合并,并不会完整地在标注中把这六项都体现出来。

需要被标注的语言可以称作对象语言(object language),即为描写者所描写、想使读者了解的语言,同时把进行标注和翻译所用的语言叫做元语言(metalanguage)。原则上说,对象语言可以是任何一种语言,对它进行标注的元语言只需是任意一种与它不同的语言或方言即可,否则便没有必要进行标注。一般而言,元语言一般使用的是文章的写作语言。比如汉语写作者可以使用汉语普通话来标注汉语方言或者少数民族语言,也可以用它来标注其他外语。

需要标注的语言成分大致可以分成两类:词汇性的成分和语法性的成分(functional elements),对象语言中词汇性的成分可以直接用元语言中的语义相同或类似的词或短语进行标注,而语法性成分,如虚词、形态等,一般使用该成分的语法类别标签的缩写进行标注。

近年来以英语作为元语言的行间标注体系发展得最为完善,但一直没有一个固定的行间标注标准,比如大家可能在缩写规则上可能并不统一,如有的人把 modality(情态)范畴缩写成 MOD,而有人把 modifier 缩写为 MOD,另外在标注的项目上、格式上、字体上各语言学家可能都存在不同。随着语法研究的深入和语言调查的需要,以及当代类型学和语料库语言学对文本标注所提出的要求,发展出了后来的莱比锡标注系统。

贰 莱比锡标注系统

莱比锡标注系统(The Leipzig Glossing Rules)由马普进化人类学研究所语言学系的 Bernard Comrie 和 Martin Haspelmath 以及莱比锡大学语言学系 Balthasar Bickel 共同制定,由10条规则和一个附录构成。10条规则主要是针对句法、语义方面的标注,附录主要是语法类别标签及其缩写形式。其实前辈语言学家在进行标注时大多都会遵循一定的标记法则,Comrie 等已经明确指出,他们所提出的标注标准反应的是传统的语法规则,在语法标注中只提出了很少的创新,并且多数是可选的。他们总结莱比锡规则的主要目的就是使已经得到广泛应用的法则和传统更加明确。标注系统目前更新到08年,并且作者指出,他们会不定期更新标注规则,所以欢迎大家的反馈。

莱比锡规则是一个以英语作为元语言进行标注的规则,它提供了三行标注的范例,即它采用了一个尽量简化的系统:第一行是对象语言的拉丁字母拼写,第二行进行逐词标注,第三行是句子的相应英语翻译。如果和上文所列出的完整的标注系统相比,莱比锡规则只标出了相应的第2、5、6三项。

对于标注规则而言,最重要的是对语素或词进行逐个标注,也就是说提供单个词和构词成分(比如前缀、后缀、中缀和词根等)的意义和语法特征。莱比锡规则明确指出,根据作者的写作目的和目标读者的背景知识,标注者可以选用不同程度的标注方式-可以标注得细一些,也可以标注得粗略一些。

莱比锡规则沿袭标注传统明确指出,它的标注单位是语素,也就是对语素进行逐个标注。但在实际标注中,对于构成复合词的词根语素不进行标注。拿汉语普通话为例,如果对“白菜”按照语素标注应该做如下第一类标注,但实际上莱比锡规则所做的是第二类标注:

③pai35ts^hai0

1 White. Vegetable

2 Chinese. cabbage

很显然构成复合词的语素并没有得到标注。

对于作为构词成分的派生性质的词缀语素(derivational affix),莱比锡规则没有明确规定,在它所举例句中,也没有看到对派生词缀给予标注的例证。

对于没有构词功能、只有语法意义的屈折词缀(inflectional affix),莱比锡规则则明确规定需要标注,并且规定了标注的方式:屈折词缀和它所附着的成分之间要用短横线“-”隔开。

更高层级的语法单位,比如大于词的短语、句子等,莱比锡规则中没有牵涉这些更大单位的方案,如果这些高层级的语法单位需要标注,需要描写者自己规定标注方案。

莱比锡规则仅仅是一个标注标准,而不是一个形态句法分析的标准框架。一种语言的形态句法系统可能存在多种分析方法,标注规则并不关注哪一种分析方法最合理,它的主要目的是在意义翻译之外,进一步对文本或例句的结构提供相关信息。当然,标注往往反映了描写者的知识背景和他对对象语言的语法系统的认识程度。一般来说,描写者的语法知识储备越丰富,对对象语言的语法体系的认识越深刻,标注也就越精确。

莱比锡规则还指出应该如何对待他人文献中语料的标注,它认为标注是语言分析的一部分,但是不是语料的组成部分,因而从一个已经出版的文献中引取例证时,如果引用者对标注有不同的分析、术语或风格,可以改变原始的标注。

莱比锡规则的十条标注规则如下:

①上下对齐

上文已经指出,莱比锡规则对例句的标注包括三项:第一行是拉丁字母拼写的对象语言,第二行是对对象语言的各成分以语素为单位进行逐个标注,第三行是英语翻译。这条规则规定,这三个项目要上下靠左对齐,并且例句行和标注行之间要以词为单位上下靠左对齐。

②语素之间要对应

标注规则规定,可切分的语素中间要用短横线“-”符号隔开,在例句行和标注行都是如此,要保证例句行和标注行有同样多的短横线。

从莱比锡规则的标注文本来看,所谓的“可切分的语素”指的是词干(stem)和屈折词缀(inflectional affix)之间要用符号“-”隔离,而当两个词根语素构成一个词时,虽然语素之间可以切分,并不使用这个符号。至于有构词作用的派生词缀,莱比锡规则没有提出明确的处理方法,从他们的标注文本看,派生词缀他们没有标注。不过既然他们已经提出标注可以根据需要有简略和详细之分,作者可以根据派生词缀在语言系统中的重要程度给予标注。

附着词或附着语素(clitics)和所附着的成分之间使用等号“=”隔开(详下)。

③语法类别标签

只有语法意义的语素通常由缩写的语法类别标签来标注,缩写字体上要采用小型大写字母(small capital)形式,缩写形式在莱比锡系统中以附录形式出现在规则之后。莱比锡规则特别提出,缩写形式可以和它提的出有所不同,比如在某种语言中,一个类别标签出现的频率非常高,更短形式的缩写则更为方便,比如可以用CPL而不是COMPL作为completive的缩写形式。同时,如果一个类别非常罕见,不必对已有的缩写标签进行再次缩写。

如果一个词可以使用缩写的语法类别标签来标注,也可以对应于元语言中的一个单词,可以选择其中任何一类,比如汉语中的作为连词的“跟”,当用英语作为元语言进行标注时,既可以使用缩写标签COM(伴随格comitative的缩写)来标注,也可以使用with来标注。

④一对多的对应

当对象语言中一个成分有几种语法功能,这几种语法功能要用多个元语言成分(单词或者是缩写形式)来标注时,这些多个功能要用圆点“·”隔开。

一对多的标注有下列几种情况:

1) 对象语言中的一个词要用元语言中的两个词来标注,这两个词之间要加“·”,比如汉语中的“出来”,相当于英语中的 come out, come 和 out 要以“come.out”的形式出现。

2) 对象语言中的一个词有多个语法功能,要用两个缩写标签来标注,这两个标签之间也应该加圆点。比如汉语中的“这些”,既是指示词,又是复数形式,所以可以使用“dem.pl”来标注。

3) 对象语言中的一个词,要用元语言中的一个词和一个语法类别标签来标注,这个词和语法标签之间也要用“·”隔开,比如法语中的 chevaux(“马”的复数形式),要使用“horse.pl”来标注。

对象语言和元语言之间一对多对应的原因很多,但是这些差异被统一应用的“·”掩盖了,所以莱比锡规则又提出了 5 条备选规则,可以显示这些一对多对应的性质,如果引入这些标注,语言成分的功能和成分之间的关系可以显示的更加具体,但是也给标注增加了难度。由于篇幅原因,此处我们不再具体介绍。

④人称代词的标注方式

这条规则是第四条规则的例外情形,即如果对象语言中一个词语同时表达人称和数,人称和数之间不必使用“·”隔开,比如河南商水方言中第一人称复数形式“俺”可直接标注为“1pl”,标示人称的“1”和标示复数的“pl”之间可以不必使用圆点隔开。这种规定也是符合莱比锡规则的标注思路:高频和显著的范畴可以有自己的专用规则。世界语言中人称代词数的区分非常普遍,因而专门为这类成分设定了一个新的标注规范。

⑤非显性成分

对象语言中的非显性成分在标注行中也可以进行标注,对这些非显性成分的标注要放在方括号之中,目标语言中这些非显性成分可以使用零形式,也可以在目标语言中标注上显性的零标记符号“Ø”,这个零标记符号和其前面的语素使用短横线“-”隔开。莱比锡标注以拉丁语中来说明这一标注,拉丁语中的 puer 可以有两种标注方式:

④puer

boy[nom.sg] (‘boy’)

或者标注为

⑤puer-Ø

boy-nom.sg (‘boy’)

英语中非第三人称单数的一般现在时动词形式就是一个零标记成分,这个零标记也可以按非显性成分进行标注。

⑥固有类别

有些固有类别在对象语言中是非显性的,但是属于对象语言的本质属性,可以在同现的其他组合中体现出来,比如法语中没有生物属性的名词其阴阳性特征是固定的,这种固有特征并不表现在名词上,而要通过和名词同现的形容词、冠词等显示出来,比如 livre(书)是阳性的,但是 pomme(苹果)是阴性的,所以说 un livre(一本书),une pomme(一个苹果),这种固定的阴阳性特征在标注项中要置于特征所属成分之后的圆括号中。

⑥ un livre

a book (F, feminine (阴性)的缩写) ('a book')

非显性成分和固有类别都是属于非显性的,但是莱比锡标注规则做了不同的处理。前者有可以与之构成语法对立的显性成分,而后者就是一个成分内在的特征,没有显性成分与之对立,这种类别只体现在它对同现成分的约束上。

④二分成分

有些语法或者词汇性成分由两个的明显的实体构成,也就是两个实体共同组成一个语言成分。典型的例证就是框式成分,莱比锡规则提出两种处理方式:

一种处理方式就是简单地重复一下,两部分都使用同样的词语或缩写做标注。

另外一种方法是用一部分来代表整体的意义,另一部分则使用语法标签标出。

比如刘丹青(2002)认为汉语中存在框式介词,框式介词是由前置词和后置词一起构成,按照莱比锡规则,可以有两种标注方式:

⑦ 在电话里(说了很久)。

tsai \ tian \ xua \ li

on telephone on

on telephone POST(postposition 的缩写)

(' (talking for a long time) on the telephone')

④中缀

莱比锡规则特意对中缀做了规定,中缀由于其位置的特殊性,和前缀、后缀使用短横线和词干成分隔开不同,在目标语言和元语言中同样都要用尖括号括出中缀。这样做使得中缀比较明显地得到辨识。当中缀有比较明显的边缘时,有时位于左侧边缘,有时位于右侧边缘,这种位置决定了在标注中中缀相对于词干的位置,当中缀靠左,在标注行中则先标出中缀,然后后跟词根。当中缀靠右,则标注中先标出词干,带尖括号的中缀后附于词干。但有时中缀没有明显的边缘,所以对于中缀的标注有待于进一步研究。

④重叠

莱比锡规则对重叠的处理和附缀相似,只是用一个波浪号“~”而不是短横线把重复的成分和词干成分隔开,同时和重复的成分相对应的地方标注出重叠的语法功能。

叁 莱比锡标注规则与汉语标注

莱比锡的标注规则对于汉语方言描写和汉语语法研究有很大作用,可以帮助不了解该方言的人更好地理解该方言的句子构造和语法体系,也帮助研究者更深入地思考该方言中各语言成分的功能。

我们 Sinotype 小组以汉语方言作为研究对象,在进行论文和专著写作的过程中,基本上以莱比锡规则作为标注的规则,同时遇到一些莱比锡规则没有明确标注或者没有处理的问题,我们对此提出了自己的观点和看法。不过语法标注是以语法分析为基础的,对同样的语言成分不同作者会有不同的分析,因而会有不同的标注方法,标注只反映了作者本人对这种语法成分的认识和分析,是可以进行修正的,正如莱比锡规则所言,引用者如果有不同的看法,作者所做的标注在引用时是可以修改的,因此标注应该尽量客观,但是考虑到对于很多语法现象并没有固定的看法,调查者和写作者的理论背景也各不相同,从这个角度而言,标注是有弹性的。

下面的一些例子是莱比锡规则在汉语中的具体运用,是 Sinotype 小组的集体讨论成果,由

于写作语言为英语,所以标注所采用的元语言也是英语。在以汉语为写作语言的论文和著作中,可以使用汉语作为元语言进行标注,这样存在的问题是需要制作一套汉语缩写的语法类别标签附在论文或专著中,所以我们可以采用综合的方式:词汇性的成分利用普通话标注,而语法类别标签的缩写仍然采用英语,可以以莱比锡标注规则中的语法标签缩写为主要依据,如果该语法项在莱比锡语法标签缩写中没有提及,作者可在文后标出该缩写的形式和所表达的意义,但是这又对写作者的英语能力提出了一定的要求。汉语标注目前还没有明确的模板和规则,还是一个值得深入探讨的问题。

3.1 格式方面的规定

我们在标注项目上,采用四行标注法,第一行为汉字,这样懂得汉字的人可以更加容易地读懂语料。第二行为汉语的拉丁字母书写,这里可以由个人灵活控制。比如我们可以采用汉语拼音的拼法,并且可以采用汉语拼音方案所规定的声调符号,但是由于方言中的很多语音和声调是汉语拼音方案所无法涵盖的,所以在此我们推荐国际音标:它可以更客观地帮助读者读出该语言,便于语音研究者观察语音规律。第三行则为遵循规则的标注行,第四行为英语翻译,使用单引号‘ ’来突出它作为翻译语言的地位。我们所采用的标注项实际上选取了完整标注的1、2或3、5、6项。这里面还有一个问题非常重要,就是汉语中比较独特的形态音位的改变,如声、韵、调的改变,有些是属于单纯的语音改变,比如连读变调,而有些变调有语法意义。根据传统的标注方法,可以单列一行标出这些形态音位改变,也就是说这一行应该标出音变前的原始读音,和前一行动实际读音进行比较可以看出哪些地方发生了变化,或者在发生改变的相应部分用黑体标出。但是这样一来标注又占据了很大空间。

根据莱比锡规则中提到的附加标准,我们认为,可以采用“\”符号来合并第3项和第4项。在国际音标一行,先标出原始读音,原始读音之后紧跟右斜线,右斜线之后是音变后的读音。在相应的标注中,如果音变有语法意义,则用相应的语法标签标出,如果只是单纯的语音改变,则在右斜线之后标注出一个 \emptyset 符号。

对语音变化,尤其是有语法意义的语音变化的标注,在莱比锡规则中只是一个可选项,没有给予过多的描述,但在汉语及汉语方言研究中特别重要,所以要加以重视。

3.2 标注中语法范畴的归并

语言中的语法范畴指的是使用固定的手段(语法结构、词序、虚词或者形态等)表达固定的语法意义。莱比锡规则虽然以标注世界上各种语言,尤其是印欧语系之外的语言为主要目的,但由于它是在印欧语言的传统研究框架下建立的,因此其语法范畴类别也集中于印欧语言的传统研究范畴。

在不同的语言/方言中语法范畴的类别不可能完全相同,比如时范畴在英语中是一个语法范畴,但是在普通话中不是。完整体/非完整体在斯拉夫语言中是一组非常重要的对立,但是英语不体现这组语法功能的对立。即使同样的语法范畴在语言中的覆盖面也可能不同,比如数范畴,在英语中覆盖到名词、人称代词和指示代词,但是在普通话中只有人称代词存在单复数对立。除了已经为人们所熟知的性、数、格、时、体、情态等语法范畴之外,语言中或方言中很有可能出现一些新的语法范畴:用固定的语法形式表达固定的语法意义,描写者要善于发现这些新的范畴。正是因为此,标注标准的语法类别标签都不可能覆盖世界上所有的语言,语言描写者应该根据目标语言的具体情况设定语法类别标签,如果语法类别标签在莱比锡规则中没有得到展示,描写者应该对这些新的语法类别加以说明。

不同语言间一般不存在功能完全对应的功能类别(虚词、形态或句法结构等),所以不同语言间相对应的语法范畴之间功能上不可能完全重合,基于此,某些印欧语言中常用的范畴常被我们弃用,我们使用了概括性更强、更符合汉语特点的范畴类别。

我们以词或结构的总体语法功能,联合词的句法位置来确立语言中的语法范畴。比如普通话中的“啊”,我们区分独立使用的“啊”(叹词)和置于句末的语气词“啊”。

我们同样区分动词后的“了”和句末的“了”,动词后的“了”我们仍然视为完整体(perfective),简称为 pfv,而句末的“了”由于它不但可以用在过去事件中,还可以用于未发生的事件中,综合这些全部功能,我们放弃了常用来描写体的 perfect(完成体)这个仅对应于句末“了”的部分功能的范畴类别,而使用概括性更强的“状态转变(change of state)”。

⑧吃饭了。

⑨快下雨了。

⑩他明天就要回去了。

如果同时放在句末和动词之后,兼有动词后“了”和句末“了”的功能,我们标记为同一个形式的语素有 2 种语法意义,2 种意义的标注之间使用圆点隔开。

同时我们还要注意,方言中的“了”或相当于“了”的成分很有可能功能上和普通话中的“了”不完全重合,要根据它们具体的语法功能进行具体分析,如果存在不同,要标注为不同的语法范畴。

再如对于商水方言中紧跟动词的“着”,由于表示“进行”“持续”和“存续”等只标示它的部分意义,并且它特殊的语法意义:强调某个对象的状态,我们据此给它一个新的名称——状态体标记(state aspect)(陈玉洁 2012)。

而有的语法意义多样,可以对应于莱比锡规则中的几个语法范畴,我们尽量采用这几个语法范畴的上位概念对其语法意义进行概括。当然,如果这几个语法意义区别性过大,也可以采用莱比锡规则中的语法范畴对其进行描写。

比如汉语中的“的”和商水方言中的“里”,它们至少涵盖了下列几种语法功能:

定语标记(attributive marker),attributive 通常指作为名词修饰语的形容词、名词、数量词和动词。

领属语标记(possessive marker),指和核心具有领有关系的修饰语标记。

关系化标记(Relative clause marker),即标记关系从句和核心之间的关系。

名词化标记(Nominalizer),即有把其他词类名词化的作用。

在具体标注中,可以按照“的”的具体语法功能对它进行标注,比如把“我的书”中的“的”标注为领属语标记,把“深奥的书”中的“的”标注为定语标记,把“她今天买的书”中的“的”标注为关系从句标记,把“我的”、“她买的”、“红的”中的“的”视为名词化标记,但这样一来割裂了“的”的各功能之间的联系,如果是对对象语言不熟悉的读者,可能看不出各种功能的“的”属于同一个“的”,并且这样标注复杂度较高,基于关联需要和简单化的需要,可以采用更加具有概括性的术语来进行概括,比如把具有连接性功能的“的”称为 ligature,而把具有名词化功能的“我的”等短语中的“的”标注为名词化标记。如果采用更加概括的看法,以上所有功能的“的”语法功能都是为了标识它所在的短语是一个名词性的短语,所以 Sinotype 小组把这种功能的“的”统一标注为名词标记(noun marker)。至于作为句末语气词的“的”,可以另外做标注。

汉语中表示动作达成的体标记(aspect marker)或阶段标记(phase marker),比如普通话中的“到”,“着”,商水方言中的“住”等,相应的英语标注为 achievement,简称为 ACH。

普通话中的“VV”或者“V 一下”形式,可标注为“暂时体”(tentative),简称为 TENTt。

经历体在莱比锡规则中也没有提及,普通话中的“过”我们使用 experience 来标注,简称为 EXP。

普通话中的“着”,可以同时表示动作持续和表示动作结束后状态持续,由于这两种功能只使用同一个标记来表示,我们称之为持续体(durative)标记。如果有方言的使用不同的语法标记来标记动作持续和动作结束后状态的持续,那就分别把表示这两种功能的体称作持续体和存续体(continuative)。

3.3 对汉语中常见的语法或词汇范畴的标注

有些范畴在汉语中功能属于比较显著,需要以缩写的类别标签对其语法意义进行标注,但是在莱比锡规则中没有注明,或者莱比锡规则对其功能的表述比较模糊,因而需要补充。

㊟格标注

通常来说,格是一种以屈折形式表现出来的名词的形态范畴,但在运用中,其意义逐渐扩展,也指以一定的句法手段,比如虚词、语序等表现出来的名词的语义范畴。在汉语中,引介名词表示伴随、处所、去向、背离等意义的介词可以标注为广义的格,比如“在”可以标注为处所格 locative,缩写为 loc,“从”可以标注为离格 ablative,缩写为 abl,汉语中标注为连词的“跟”和“和”可以标注为伴随格 comitative,简称为 com。当然,也可以采用比较粗略的标写方式,如果语言中有相应意义的词语,可以使用相应的词语进行标注,比如把“在”标注为 at 或 in,把“从”标注为 from 等。

㊟量词

莱比锡标注系统中,把量词 classifier 标注为 clf,但是在对汉语进行标注的大部分论文和专著中,都把量词标注为 cl,这个处理节约了空间,基于量词在汉语中特别显赫的特征,我们采用 cl 来标注名量词。同时使用 vcl 来标注动量词,因为动量词和名量词虽然都起量化作用,但在汉语中有不同的语法意义:名量词是对指称对象做量化,而动量词则是对事件或动作的量化,所以特意标注出来,以表示区别。

㊟有多个意义的成分的标注

对于有多重语法功能的成分,如给,可以采用词汇意义和语法意义相结合的形式进行标注,比如给,可以标注为 give,然后使用下标的小型大写字母标出他们具体的语法意义,如 benefactive, comitative, malefactive 等等。

指示词也是这种词汇意义和语法意义的结合体,也可以采用这个下标形式进行标注,比如如果语言或方言中仅有两个指示词,一个指远,一个指近,远的那个标注为 dem_{dist}, dem 是 demonstrative(指示词)的简写形式, dist 是 distal 的简写形式,因为指示词和远指不属于两种不同的语法功能,远指只是说明了指示词内在的语义特征,所以以下标形式标注在 dem 之后。同样,近指指示词处理为 demprox, prox 是近(proximal)的简写。不过也可以直接标注为 that、this。有的方言中还会出现中性指示词,需要标注为 demneut, neut 是中性(neutral)的缩写形式。

㊟人称代词

人称代词要标示出人称,用 1, 2, 3 代表,同时如果存在数的对立,需要标示出数特征,我们遵循莱比锡规则,人称和数特征之间不用圆点隔开,比如第一人称单数,标注为 1SG。

⑤对于语音改变的标注

莱比锡规则所制定的可选规则中,有一条和汉语密切相关,需要我们特别加以注意。它指出,如果在目标语言中一个语法特征通过形态音位改变(如元音交替(ablaut)、元音变化(mutation)或声调改变)来实现,在标注行中要使用斜线“\”把这些形态音位改变所带来的语法意义和它们所附着的成分本身的意义分隔开来。并且举德语为例:

- ⑪ unser-n väter-n
our-DAT.PL father\PL-DAT.PL ('to our fathers')

复数形式 väter 来源于单数形式 vater。

汉语中通过变调表达语法意义的情况很多,比如商水方言中,指示词“那”作为名词化标记时一定要读作轻声,轻声变调此时有语法意义,可以标注为:

- ⑫我夜个买那
uo ɿ ie ɿ · kə mai ɿ · na ɿ
1sg yesterday buy DEM_{dist}\NM ('those I bought yesterday')

再如河南浚县方言中(辛永芬 2006),可以用动词变韵来表示完整体(perfective)的语法意义,可以标注为:

- ⑬买一斤盐
mai ɿ \mɛ ɿ i ɿ tɕin ɿ ian ɿ
buy\PFV one CL salt ('(Someone) bought a catty of salt.')

⑥词缀的标注

从句法功能上说,汉语中的词缀可以分为两类:一类是有构词作用的派生词缀,派生词缀除了具备一定的语法意义之外,还有构词功能,可以改变词性或构造新词;另一类仅有语法意义,可以附着在单个词之上,也可以附着在短语之上。

莱比锡规则提出,可切分的语素(segmentable morpheme)之间要用“-”隔开,但在实际语料的处理中,他们没有使用短横线来隔开复合词和派生词中的各个语素,短横线只用于隔开词干(stem)和只起语法作用,没有构词功能的屈折词缀(affixed inflection,汉语语法学称为构形词缀)。构词的派生词缀(derivational affix)由于和其所附成分构成一个词,这个词一般可用元语言中的词或短语进行翻译,所以可以不必标注,这种做法有一定道理,因为如果对词内成分进行标注,会造成标注过于繁琐复杂。但对于汉语来说,因为以前对汉语的形态注意不够,方言语法中的构词词缀需要得到描写和标示,对这些词缀成分的标注有助于帮助读者更进一步地了解对象语言的构词法以及词缀所体现出来的语法特征,从研究角度而言有重要作用。这个可以根据作者的需要灵活选择。

我们只标注只有语法功能却没有构词功能的屈折词缀或者叫构形词缀。按照这种观点,汉语中前缀“阿”、“老”,后缀“子”、“儿”、“头”、“化”等都不必进行标注,因为这类构词成分和所附词根一起构成语言中的一个词,可以用元语言中相应的词或语素进行翻译。但是在标注中有一些特别现象需要注意,比如对普通话中“儿”的标注。虽然很多语言学家(吕叔湘 1982;赵元任 1968)把“儿”看做和“子”、“头”具有同样地位的构词词缀,这种观点也是现代汉语语法研究中比较通行的观点,但也有部分语言学家,如高名凯(1986)、刘雪春(2003)认为,普通话中的“儿”实际上可以分为两个不同的语素:一个是构词成分,一个是构形成分。刘雪春指出,“儿”在词中有三种作用:第一是把一个现成的词变成另一个词,如“嘴-嘴儿”;第二

是把不成词的语素或语素组变成词,如“褂-褂儿”,增加或改变词的附加意义,如“碗-碗儿”。前两种功能体现了“儿”是比较典型的构词成分,后一种功能中“儿”不起构词作用,而主要起小称以及由此发展而来的正面积积极意义的语法功能,所以是一个构形后缀,或者叫屈折后缀,这部分“儿”在标注中需要加以标注。

⑭ 盖子

kai ㄋ · tsi

cover (‘cap, cover’)

⑮ 碗儿

uan ㄣ · er

bowl-DIM (‘bowl(with diminutive meaning)’)

④在汉语方言中还有些融合的形式,比如商水方言中的[iuo ㄣ],是数词“一”和量词“个”语音融合的结果,这样的融合形式要使用两个相对应的意义把它标注出来,这两个意义之间用圆点隔开,比如如果以英语为元语言进行标注,可以把商水方言中[iuo ㄣ]标注为“one.CL”使用汉语为元语言,则标注为“一.量词”。

㊸ 重叠

重叠是汉语中常用的一种语法手段,在方言中这种语法手段也得到了广泛应用,所以我们特意关注对有语法意义的重叠现象的标注。

汉语中的重叠可以分为三类,第一类重叠构成多音节语素,比如“蝈蝈”、“猩猩”、“奶奶”等,这部分重叠属于音节的重叠,其中非重叠的单音节形式不构成语素。第二类重叠是一种构词手段,比如“姐姐”、“常常”,这部分重叠形式中非重叠的单音节形式也是自由语素(即可以视为词),重叠之后仍然构成一个词,重叠前后词的意义并没有改变。第三类重叠是一种具有语法意义的重叠,自由语素重叠之后有了新的语法意义,比如形容词、动词、量词和部分名词的重叠。赵元任(吕:105)引 Bloomfield 的观点认为,重叠可以视为一种变化,也可以视为是一种词缀。莱比锡标注规则中,把重叠作了不同于词缀的特殊处理,词缀使用“-”隔开,但是重叠使用“~”符号隔开,我们按照莱比锡规则,把重叠视为一种特别的形态变化,使用专用符号“~”把基本形式和重叠形式隔开。赵元任没有区分汉语中这三类重叠,由于重叠前的形式不是语素,所以第一类重叠不需要标注,第二、三类涉及到构词法和句法的重叠需要标注。

⑯ 姐姐

tɕie ㄣ ~ · tɕie

elder.sister ~ rdp

第三类有语法意义的重叠中,除要用“~”符号标注出重叠外,还要标出它的语法功能。

⑰ 走走

tsou ㄣ ~ · tsou

walk ~ rdp.dlm(dlm 是 dilimitative 短时体的缩写)

但同时赵元任已经指出,这种重叠形式和普通词缀之间存在很大不同,它没有固定形式,采用它所附着的形式或所附着的形式的一部分。

汉语中有许多种类型的重叠,我们只关注有语法意义的重叠现象,而仅仅是构词的重叠,比如“爸爸”、“妈妈”等,不再标出。和莱比锡规则不同,我们区分重叠的基本式和重叠部分,在基本式和重叠部分之间使用波浪符号隔开,重叠部分我们认为有两个功能,其一它是一个有

粘着功能的附缀性成分,第二重叠有一定的语法功能,因而我们使用缩写的语法标签 REDP 显示出它是重叠部分,然后加圆点,圆点后标出重叠后的语法意义。

④关于句末语气词的处理

汉语中的句末语气词有表示情态和式的意义,比如普通话中的句末“啊”可以表示感叹语气,河南商水方言中的句末语气词“着”有可能意义。情态和语气的关系非常复杂,比如商水方言中的“着”,和句末“了”、“吧”等一样有完句功能,同时又有可能意义(陈玉洁 2012),我们不区分情态和语气。如果把他们看成独立的词,那么和其他成分之间可以直接使用空格隔开。至于标注标签,可以认为他们属于情态类成分,标注为 MOD(modality 的缩写),当然如果不认可这种分析,也可以标注为 PRT(Particle 的缩写)。同时我们必须认识到,这种句末语气词有非常强的依附性,不能单独成句,必须依附在其他有词汇意义的语言成分上。赵元任(1968)曾把这类附着在句子或者短语上的成分,可以称为 clitics,我们翻译为附着词,中间使用“=”分开。Loos 等人把附着词(clitics)定义为:附着词有单词的句法特征,但是语音上往往附着于其他词类成分。这类词的特征如下:

- 1) 语音上附着但是句法上自由,即可以独立使用,句法上可以分析为一个词。
- 2) 作用于词组或者句子层面。
- 3) 如果不附着于其他成分,不能出现在正式的话语中。
- 4) 常常只有语法意义,但是没有词汇意义。
- 5) 和代词、介词、连词、助动词等类似,属于封闭性的词类。
- 6) 附着于词语的最外层,在派生词缀和屈折词缀的外层。
- 7) 常常可以附着于多个范畴的词类上,比如核心名词,非核心名词,介词,动词,副词等词类上。
- 8) 语音上是非重读的。

从这些标准来判断,汉语中具有完句功能、附着于句末的被称作语气词的成分,全部符合附着词的特征,所以是非常典型的附着词。由于它所作用的句法范围是整个句子,所以是附着在整个句子之上的,但语音上它常常和它前面的成分合并,形成新的语气词。汉语中的语气词我们在其前加等号显示它的附着词地位。

⑤关于汉语中补语的标注

汉语中的动补结构是汉语中已经高度语法化结构,这类特别的结构在标注中需要加以特别说明。莱比锡标注规则由于基于西方语言的语法规则而建立,对此结构没有给予说明。这显示了莱比锡标注规则在面对具体语言的标注时确实要根据语言事实进行补充。

我们根据语法化程度的不同把汉语中的补语分为两种类型进行标注。

语法化程度高的动补结构,我们把结构中各成分之间使用“+”连接,这些语法化程度高的动补结构包括:

- 1) 能性补语结构。指用动补结构表达可能意义的结构。可能意义无法直接从动词和补语的相加中得到,是结构本身所带来的意义,因而此结构间各成分的关系非常密切,使用“+”符号可以看出这种结构本身的高度语法化。在能性补语结构中,如果带有语法标记,则该标记标注为 potential marker,简称为 POT.这个语法标记有可能和通用补语标记重合,如普通话中的“得”,这样造成该语言或方言中的动补结构出现歧义,而在有的方言中,该语法标记和通用补语标记不同,比如河南商水方言。能性补语的否定形式同样使用“+”连接,只是否定标记兼



做可能标记。这一点可以从下文中通过和结果补语的比较看出。

普通话

⑮洗得干净

çi ʌ + tə ʌ + kan ʌ tçiŋ ʌ

洗 + pot + 干净 (‘能洗干净’)

⑯洗不干净

çi ʌ + pu ʌ + kan ʌ tçiŋ ʌ

洗 + neg.pot + 干净 (‘不能洗干净’)

⑰洗得不干净

çi ʌ tə ʌ pu ʌ kan ʌ tçiŋ ʌ

洗 VCM NEG 干净 (‘洗得不干净’)

商水方言:

⑱洗干净了

çi ʌ + kan ʌ · tçiŋ - liao

wash + clean + pot (‘can wash (it) clean’)

⑲洗不干净

çi ʌ + pu ʌ + kan ʌ · tçiŋ

wash + neg.pot + kan ʌ tçiŋ ʌ (‘cannot wash (it) clean’)

2) 其他使用补语形式的结构。如果在方言中,固定的动补形式(加标记或者不加标记)用来表示某一专门的语法意义,这一专门的语法意义是动补成分相加所无法得到的,这个结构的各成分之间要使用“+”连接,以显示结构间各成分的密切关系。

3) 需要使用“+”连接的结构是不使用标记的动补结构,当然这类结构有可能有特殊语法意义,比如方言中有可能使用“洗干净”来表示可能意义,也有可能是一个简单的动补结构,无论如何,这部分结构已经有了一定程度的词汇化,赵元任把它们归为复合词范畴,我们使用“+”对此进行标记。为了处理方便,我们把不使用专门补语标记的趋向补语也归入此类。

⑳吃饱

eat + full (‘to be full’)

趋向补语的语义状况比较复杂。有的是意义比较实在的补语,标示运动的方向和路径,这部分看做补语,而有的已经虚化为体标记,这部分体标记由于是附着在动词之上的,根据我们对体标记的标注方式,这部分趋向动词不看做补语,而视为体标记,和所附着的主要动词之间使用后缀“-”形式隔开。

肆 结论

标注牵涉到对语法系统的选择、认识和处理,根据现在语法研究的现状,尤其是功能词和固定结构的语法功能到现在很多问题还没有定论,细节问题各家处理还存在很大不同,因此标注规则具有一定的灵活性,个人所采用的标注体系和自己对语法的了解和语法观念密切相关,不可能完全统一。标注能清晰地梳理语言中各个单词或语素的用法,提高了语料的精确度。



参考文献

- 高名凯 1984/1986 《汉语语法论》,(北京)商务印书馆
- 刘丹青 2002 汉语中的框式介词,《当代语言学》第4期
- 刘雪春 2003 儿化的语言性质,《语言文字应用》第3期
- 吕叔湘 1956/1982 《中国文法要略》,(北京)商务印书馆
- 辛永芬 2006 河南浚县方言的动词变韵,《中国语文》第1期
- 赵元任 1968/1979 《汉语口语语法》,(北京)商务印书馆
- Comrie, Bernard & Martin Haspelmath & Balthasar Bickel 2008 Leipzig Glossing Rules. <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>.
- Loos, Eugene E. etc. Glossary of linguistic terms. <http://www.sil.org/linguistics/GlossaryOfLinguisticTerms/contents.htm>.

The Leipzig Glossing Rules and Its Application in the Research of Chinese Grammar

CHEN Yujie, Hilario de Sousa, WANG Jian,

NI Xingxing, LI Xuping, CHEN Weirong & Hilary Chappell

Abstract Interlinear gloss, as a useful device for linguistic description, is often applied by linguists to facilitate classification, preservation and analysis of language data. It is also widely used in the realms of modern linguistics, i.e. linguistic typology and corpus linguistics, and even in computer sciences as Cathy Bow has done in his endeavor to work out software and procedures to better off the generation and processing of contexts. Unfortunately enough, interlinear gloss is still far less employed than necessary in researches of Chinese study. This paper introduces the Leipzig glossing rules and its application in the annotations of Chinese grammar with the cases of language description by the Sinotype project of EHESS.

Key words interlinear gloss, the Leipzig gloss, annotation of Chinese grammar

本所三种期刊均入选“2013 中国最具国际影响力学术期刊”

据《中国新闻出版报》2013年12月30日报道,由中国学术期刊(光盘版)电子杂志社、清华大学图书馆、中国学术文献国际评价研究中心遴选的“2013 中国最具国际影响力学术期刊”名单公布,包括自然科学与工程技术类期刊175种和人文社会科学期刊56种,各根据“国际影响力指数CI”排序。语言学期刊入选6种,分别是(顺序按在原名单中的先后):《中国语文》、《世界汉语教学》、《汉语学习》、《外语教学与研究》、《方言》、《当代语言学》。在同时公布的“2013 中国国际影响力优秀学术期刊”名单中,有《语言教学与研究》、《外语界》、《语言研究》和《外国语(上海外国语大学学报)》4种语言学期刊入选。