

文本标注的历史性跨越与改进空间

——以“中国民族语言语法标注文本丛书”为例*

刘丹青

[摘要] 本文指出在中国民族语言记录描写中，文本的语法标注是一个薄弱环节，而“中国民族语言语法标注文本丛书”采用基于类型学语言多样性而设计的多行标注法，更充分、更精准地提供了句子结构的句法语义信息，显著提升了文本的语言学价值，实现了文本标注的历史性跨越。文章以具体实例的细致分析，既展示了新型标注法的6方面优点，也指出了丛书作为新的尝试尚存的4方面问题，为完善民族语言标注法提出了建议。

[关键词] 民族语言 文本标注 逐词标注 语法素

引言

现代意义上的中国少数民族语言调查记录始于1920年代，迄今已一个多世纪。中华人民共和国成立后，这一事业进入全面快速发展期，涌现出大量优秀成果（孙宏开等 2007:3）。

对少数民族语言本体的介绍，主要以两种形态呈现。一种是语言学者的描写，包括语音、词汇、语法的专书系统描写或论文专题描写，都有大量例句；另一种是调查语料的直接呈现，包括专著后附的长篇语料和单出的文本语料集，如《壮侗语族语言文学资料集》（1983）。两种文献的语料出现于中文书刊时，都有音标并带汉语注译，便于读者既知其音，又明其义。

总体上民族语言语料能满足读音方面的基本需求，也大致能实现对整句意义的理解。但是，在语法结构方面，不同的注译方法能提供的信息数量和质量高下显著。影响因素有切分细度、语法框架、语种类型特点、标注的精确度等。有些语料只提供整句翻译，读者难以明了句子内部结构如何；大部分语料或多或少提供词语注释，但由于虚词、形态等方面注释缺乏统一规范，很多分析和标注受汉语眼光和术语系统的限制，提供的语法信息用途有限。

我们面对的问题，在其他语言包括西方学者所记录的各大洲语言材料上也会遇到。只受过西方传统语言学训练的学者，也常以印欧语的眼光标注类型迥异的语言，如早期传教士对汉语的记录。对此，面向全人类语言、重视多样性的语言类型学，在大量调查研究和跨语言比较基础上不断总结经验，逐步建立起既能尽量反映语言个性、又有相对一致的原则、便于互相理解和引用的标注规范。这类规范以多行标注的莱比锡标注法为代表，已得到语言调查

* 本研究获国家社科基金特别委托项目“中国大百科全书第三版”（21@ZH015）子项目语言文字卷、深圳市孔雀鹏城学者启动项目支持。初稿曾在“中国民族语言学会语言类型学专业委员会第五届学术年会”（玉溪师范学院 2022.11）线上宣读。

者的广泛采用，成为当代语料标注的国际通行规范（黄成龙 2005；陈玉洁等 2014；罗天华 2023）。另一方面，莱比锡标注的元语言是英语，而中国出版的语言学文献主要以中文呈现。因此，陈玉洁等（2014）也对莱比锡标注的中文化处理作了提议和讨论，以便于汉语学界在中文出版物中吸收莱比锡标注的长处，也便于这些资料译成外语、走向世界。

针对民族语言学界比汉语学界更为复杂的情况，江荻、黄行、燕海雄等学者组成的团队，承担了“中国民族语言语法标注文本丛书”（下文简称“丛书”）重大项目，引领前沿，采用与莱比锡标注法类似的多行标注法，在民族语言学界跨出了重要一步，以此为数十种民族语言提供更高质量的标注语料，现已出版 20 余种。这一项目不但在真实语篇的容量上大上台阶、贡献甚巨，而且在标注质量上实现了历史性提升甚至飞跃，由此涌现出大量高质量、可引用的真实语料，为民族语言学、语言类型学、历史比较语言学、接触语言学、语篇语言学等学科提供了高价值宝贵资料。我们为这一重大成果鼓掌点赞并期待其陆续有来！

作为历史性跨越的第一步，涉及数十种类型、谱系和研究基础差异都很大的语言，丛书的标注也存在一些可理解的欠成熟之处。本文拟以这套丛书为例，简要分析这种新型标注法带来的语料价值提升，同时讨论存在的一些问题，促进这类工作的优化^①。

一 丛书文本标注法的进步及其意义

（一）逐词逐语法素标注

以往的民族语言文本资料，标注所切分的单位大小不一，有些只提供整句翻译，无法将语义落实到词；也有很多材料有词级标注，但是主要是实词的对译，虚词和形态要素的标注做法不一，信息常失之模糊，主要提供一些传统语法范围的尤其是汉语界常用词类的标注。这种标注，大致可以概括为“对译+词性”的标注。我们知道，对于认识句子结构来说，虚词和形态提供的语法信息比实词更重要。以往标注法的主要弱点就在虚词和形态要素的标注上，空泛的虚词词性如“助词”之类信息有限。况且，汉语是一种形态很少的语言，形态在语法学中权重很轻；而很多民族语言的形态比汉语丰富复杂，如阿尔泰语系语言、羌语支和藏语支等藏缅语言，对这些语言的标注多受汉语语法学影响，使得形态标注成为薄弱环节。

莱比锡标注法要求按词或按语素标注。实词和虚词都要标注到词，词内语素只标注构形形态要素，构词词缀因为直接影响句子结构而一般不标注（陈玉洁等 2014）。有些黏着语和多式综合语一个词内可叠加多个构形成分，需逐个标注；有些屈折语一个构形成分包含多种语法范畴意义，需分别标注并用墨点隔开。虚词和形态成分在类型学界获得一个合称叫“语法素（gram）”（Haspelmath 2003）。因此，标注的切分原则可以表述为“逐词逐语法素标注”。

这套丛书总体上实现了逐词逐语法素的标注，语法信息质量显著提升。例如藏语甘孜话（燕海雄、江荻 2016:84）：

(1) dza³¹mo⁵⁵ yø³¹ ta⁵³re⁵⁵ tsha⁵³tei⁵⁵ dza³³bø⁵³ le³¹ tche⁵⁵re³¹ se⁵⁵ni³¹ dza⁵³la⁵⁵ re³¹.

王后 AGT 看见 PFV 以后 国王 ALL 什么 COP 问 SEQ 说 IPFV COP

皇后见此情景，就去探问国王的心思，问国王为何不高兴。

例(1)中，每个实词、虚词和形态要素都得到了清晰标注。AGT 为施事格，是藏缅语

^① 由于笔者尚未通读丛书已经出版的卷册，本文所谈只能是举例性的，没有量化评估的作用。

中常见的凸显施事功能的要素，但是该成分在汉语中没有对应物（与话题标记性质的语气词功能迥异），传统的汉译式标注有时就标为功能不明的“助词”，甚或不标注。PFV是完成体，与靠近句末的IPFV（未完成体）相对。汉语完成体一般用体助词“了”表示，而非完成体没有专用标记，传统标注方法难以对译。ALL是向格（方向终点标记，该多功能语素也能表位格，即处所标记）。两处COP体现了藏语系词的多功能性，这种句子在汉语中不用系词，若用“是”来对译，反显别扭。SEQ是顺时连词，其用途与任何汉语虚词都难对应，因此用这个功能标记比生硬翻译或不标注都强。此例展示新型标注法确实更加简洁清晰地提供了充足的语音、词汇和语法信息，让读者明白整句译文之义是如何通过句法成分和结构来实现的。

（二）形态标注

世界上多数语言形态多于汉语，构形形态信息是句子语法信息的重要部分。但是，长期影响民族语言学研究的汉语语法学，却因为自身分析性语言的特点而不需要对形态多加关注，导致用中文术语标注形态会捉襟见肘。现在引入新型标注法，并在书前设语法范畴缩写的术语对照表，标注行就能对形态要素充分标注。如独龙语的例子（杨将领 2020:172）：

(2)	tə ³¹ mu ⁵³	mu ²⁵⁵	pəŋ ³¹	tɔ ²⁵⁵ tɔ ²⁵⁵	ɛw ³¹	pə ³¹ tei ²⁵⁵ teəl ⁵³	xw:p ⁵⁵	wa ³¹
	tə ³¹ mu ⁵³	mu ²⁵⁵	pəŋ ³¹	tɔ ²⁵⁵ -red	ɛw ³¹	pə ³¹ tei ²⁵⁵ teəl ⁵³	xw:p ⁵⁵ -3p	wa ³¹

鹰 天 ABL 俯冲-RED RFLX 小鸟 抓-3pe HRS

鹰从天空中俯冲下来抓住了小鸟。

杨著和丛书部分卷册一样，原文行下增加一行语法分析，对原文添加一些语法分析。例(2)只是独龙语故事中一个不长的分句，但是作者提供了5项形态标记。ABL是从格标记，表示源点。RED代表重叠，标明第二个tɔ²⁵⁵是第一个tɔ²⁵⁵重叠的产物，与第二行的-red对应。在传统的中文标记法中，重叠或用汉语的动词重叠形式来标注。但这并不妥，因为汉语动词重叠的语义功能难以代表其他语言中的动词重叠。RFLX是反身语素，代表独龙语的反身—中动态标记，若用汉语“自己”来对译，在此并不自然，而且“自己”也有其他语言反身语素所不拥有的功能，与独龙语的该标记并不完全对应。所以，这里用范畴标记RFLX来标注，能客观反映独龙语反身语素的特色用途。-3pe对应分析行中的-3p，这是汉语无法直译的范畴，表示动词的人称一致关系——第三人称的强调式，而且该形式还是单复数不分的形式（独龙语另有单数复数有别的一致关系形态）。这更是难以用目标语言标示的语法意义。以往有些文献遇到这种内容就干脆不标，因为反正也不会影响到句义的理解，但是它会妨碍对句子结构的理解。最后的HRS更属汉语没有的范畴——动词的示证形态，标示信息来源中的听说类，即非言者亲见，而是听来的信息。这么多语法信息就在这一行标注中得到了完整清晰的体现，体现了新型标注法的强大能量。这在传统的“对译+词性”标注法中是很难实现的。

（三）虚词标注

虚词是任何语言中都存在的语法手段，在分析性语言中尤其重要，因为这些语言形态稀少，很多语义要素和语法关系通过虚词手段表示。在民族语言学的传统标注中，虚词标注相对薄弱，很多功能各异的虚词只被归为有限的若干词类或小类进行标注，参照的往往还是汉语语法，故大量的虚词信息在转写和标注中流失或打折。丛书在这方面进步显著，值得称道。

下面以黎语白沙话一个虚词为例。黎语属壮侗语，其分析性类同汉语，形态很少，但是有一些重要而常用的虚词跟汉语难以对应，在民族语言的传统引例和语料中很难得到充分而合理的标注。丛书在这类虚词上的标注展示了显著的优越性。如白沙话的na³³，是基本远指

指示词，相当于“那”，与近指代词 *nen*³³ 相对，并且论元功能和限定功能（在 NP 之后）都有；还可以作远指处所词，相当于“那里”，例如（吴艳 2021:95、222、24）：

(3) na³³ vai³¹ don⁵⁵ nə¹¹ tham¹¹ lo¹¹. 那肯定不是他老婆了。

那 不是 老婆 3sg-他 MOOD IND

(4) lan¹¹ tuk¹¹ te¹¹ na³³ ka³¹ uap¹¹ dak¹¹ uap¹¹ lan¹¹ tuk¹¹ ta³¹ məu⁵⁵ na³³.

CL-个 孤儿 那 也 喜欢 真 喜欢 CL-个 彩虹姑娘 那

孤儿也特别喜欢那个彩虹姑娘。

(5) la:i³³ ko⁵⁵ nə¹¹ fat⁵⁵ rəu¹¹ khəu³³ khur³³ na³³. 看见她哥哥正在那里休息。

看见 哥哥 3sg 正在 休息 PREP 那里

na³³ 在例（3）中充当主语，代指那个人，例（4）中充当主语和宾语的限定语，例（5）中充当前置词的处所宾语。这些都符合指示代词的功能。作者也是用“那”“那里”来标注的。但 na³³ 的功能不止于此，吴艳（2021:34）认定该词指示代词以外的功能有 5 项：定指标记、话题标记、连词用法、标句词、语气词。我们着重说一下与指示代词关系密切的前两项。

定指标记，书中以缩写 DEF 标注。在国内的语法学传统中，基本没有“定指标记”这一概念。它对应部分印欧语的定冠词。一般认为东亚语言没有定冠词，况且指示词本身也有定指作用，因此，定指功能很可能就用汉语指示词“这”“那”来标注了。但这样的标注，远不能反映已经成虚词的定指标记的功能。吴艳（2021）正是凭借定指标记 DEF 的标注，才让我们充分体会黎语这个重要虚词（相当于后置冠词）的显赫功能。

首先，白沙话的定指标记虽源自指示代词，但该功能已不是指示代词所能覆盖的。它所标注的文本中的有定名词语，在对译的汉语句子中大部分未见指示代词。其间的最大差别是：指示代词以直指（deixis）为原型功能，有时兼有回指（anaphora）功能，特别是在没有定冠词等定指标记的语言中；而定指标记则以回指为原型功能，有时也能借助语境起直指的作用。白沙话 na³³ 的定指功能已经与指示代词用法有所分化，最明确的证据是可以与指示代词连用，各司其职——指代和有定标记，如例（6）（7）中标注为 DEF 的成分（吴艳 2021:34）：

(6) nen³³ na³³ pla³¹? 这是什么？

这 DEF 什么

(7) na³³ ti³¹ tsu³¹ duan⁵¹ nen³³ na³³ thom³¹ na³¹. 那这段故事就是这样子。

那 是 一 CL 这 DEF 这样 （引者：字面意思即“那是这样的这一段”）

这几例中的定指标记 na³³ 绝不能译为“那”，它甚至超出了英语定冠词的使用范围。

除了这些突出的例子，还有大量的 na³³ 加在非冠词语言无须加指示词的位置，甚至加在冠词语言也不加定冠词的位置，只是为了标注有定指称，如（吴艳 2021:79、75、77）：

(8) rip⁵⁵ za³¹ na³³ mai¹¹ khai³³ pai³³ don⁵⁵ nə¹¹ na³³ mou¹¹ tsu³¹ ton³¹ fei¹¹.

蜈蚣蛇 DEF 就 CAUP-叫 妻子 3sg-它 DEF 烧 一 CL-堆 火

蜈蚣蛇叫它老婆烧一堆火。

(9) pai³³-duai¹¹ zi¹¹ na³³ ka³¹ ai³³ da:m¹¹ phə¹¹. 二女儿也不想嫁。

F-第二 DEF 也 不愿 想 嫁

(10) za³¹ na³³ təm³³ fei⁵⁵, təm³³ vou⁵¹, ai³³ phə¹¹ mə³¹ me¹¹ mə³¹.

蛇 DEF 不懂 走 不懂 跑 不愿 嫁 IMPM 妈妈 IMPM

蛇不会走不会跑，我不想嫁啊，妈妈呀。

例(8)第二个na³³用于本身已经是有的领属结构(它老婆)。例(9)是家庭语境下的排行名词(二女儿),也带定指标记。例(10)“蛇”是类指性的,在冠词语言中常用定冠词表达。这几例的情况在汉语这样的非冠词语言中都不宜用指示词;而在英语中,“序数词+名词”结构需要带定冠词(如:the first son),“名词+基数词”的序数结构则不带定冠词(如:Volume Two),带领属语的名词也不带定冠词(如:the snake's wife)。na³³虽称为定指标记,但句法属性上属于“虚词”而非形态,因为它可以加在短语上,并有稳定的音节地位。

定指成分有较强的话题性,白沙话的na³³也发展出了话题标记的功能,有些还处在兼有定指功能和话题标记功能的中间状态,作者标以DEF/TOP,如(吴艳 2021:101):

- (11) su³³saj¹¹ na³³ mə¹¹ lei³¹pla³¹ kha¹¹ nen³³? 以前你为什么来这里?
从前 DEF/TOP 2sg-你 干什么 上 这里

这样的标注,既说明了功能项之间的共时关联,也提示了它们的历时发展轨迹。按传统,对该词的标注大致会有如下几种:①用“那”对译,但其很多功能无法译成“那”,尤其是用在表“这”的近指指示词后时;②用词类名“指示代词”标注,但是该词多数用法都是回指,而非直指,大部分用例都难以译成汉语的指示词;③用“助词”类标签,这除了提示它是一个虚词,不含任何语法语义信息。而现在标注为DEF,让读者了解到该虚词的主要功能已经是回指而不是直指,其搭配范围之广、出现频率之高、使用强制性之严格,都远超一般指示词,表明定指是白沙话中高度语法化的范畴,甚至是显赫范畴,在同区域语言中非常突出。

从书中定指标记的标注不限于黎语。藏语甘孜话中,来自指示代词的高频定指标记te³¹也以DEF标注,展示了定指范畴的显赫地位。例如(燕海雄、江荻 2016:85):

- (12) nə¹¹ ta⁵⁵ ma³¹tu⁵⁵ su¹⁴ ma³¹ khe⁵⁵ yə³¹ a⁵⁵te⁵⁵ te³¹ ta⁵⁵ te³¹ ma³¹tse⁵⁵ru³¹tse⁵⁵ re³¹.
如果 马 只有 牛 H-NEG NOM GEN 乞丐 DEF 马 DEF 不得不交 COP
那个只有马没有牛的乞丐,就不得不把马给我们送来了。

总之,丛书超越传统,凸显范畴,显著提升了虚词标注的精确性、通用性和简洁性。

(四)语义精准性

传统的“对译+词性”标注法,实践中往往优先采用对译法。由于语言各有自己的词汇、语法库藏结构,很多词在目标语中不具备严格对应的词。在对应性不强的情况下,对译法往往难以准确传递句法语义信息,尤其对虚词或半虚化词语而言,进而导致认识的偏差。如前文所举黎语白沙话na³³。如果简单地用“那”对译,就不能充分体现其定指功能及其他附带功能。用DEF“定指标记”,则准确反映了该虚词的基本指称功能,提升了语义精准度。

人称代词也有很多“雷点”,看似简单的对译,隐藏着不少误差。如民族语言中的第三人称单数代词,常对译为“他”,或根据语境所指对译为“她”“它”。然而,书面汉语中,“他”“她”有性的对立,字形上与英语he、she对应。实际上东亚地区几乎所有语言和汉语所有方言都不能区别第三人的性,因此,用其中的任何一词来对译民族语言的第三人称单数代词,都会造成信息溢出,无端增加原词没有的信息。现在,新型标注法用缩写代码来标注人称代词,东亚地区语言中的第三人称代词实际上大都是3sg,即第三人称单数,语境中指男、指女的解读不是代词本身的词义。所以,这个代码比用“他/她”和“he/she”对译要准确。如黎语白沙话的代词nə¹¹,不管在语境中所指性别是什么,吴艳(2021)都标注为3sg。其他代词亦然。再如普通话口语有普通的“你”和尊称“您”的差别,而有些民族语言第二人称代词没有尊称/非尊称的区别,因此无论用“你”或“您”对译都不准确,信息溢出,而用

2sg 来标注就无此问题。其他还有双数与复数的分合、复数代词包括式与排除式的分合，都难以通过代词对译来圆满标注；最适合用人称、数标注，需要时加上包括式或排除式等标签。

丛书在这方面改进甚多。如体标记不用“了、着、过”对译，而用 PEF（已行体）、PFV（完成体；完整体）、PROG（进行体）、CONT（持续体）等标注。不同语言的体标记之间，高度对应的情况实属罕见，直接用具体的汉语体标记来对译，难免牵强，还有一些体貌在汉语中根本没有体标记可对应，不如用范畴代码稳妥。至于每个体标记的适用语境和功能个性，可以通过对比译文并观察大量文本获得，这正是文本语料的优势所在。有些虚实之间的词，如系词，也不一定要用具体词项如“是”来对译。不同语言的系词功能差别很大，如印欧语形容词作谓语要用系词，而汉语、壮侗语等形容词可以直接作谓语。有些语言系词的适用面比汉语大得多，如藏语甘孜话，很多句式以系词煞尾，其中多数无法译为汉语“是”，而用系词的代码 COP 来标注，就自然得多。例如（燕海雄、江荻 2016:62、70-71）：

- (13) na⁵⁵ ña⁵³mu⁵⁵ ña⁵³mu⁵⁵ ñe⁵⁵ne³¹ dza³³bɔ⁵⁵ tsi⁵³ jo¹⁴ re³¹, khe⁵³ mə³¹se⁵⁵ tei³¹ lə¹¹
很 早 早 以前 国王 一 有 COP 3sg-AGT 人民 PL DAT
di³¹wu¹⁴zɔ³¹ la⁵⁵ re³¹. 很久以前，有一个国王，他对人民残酷。
残酷 IPFV COP
- (14) na⁵⁵ñɑ⁵³mu⁵⁵ ñe⁵³pa⁵⁵ tsi⁵³ jo¹⁴ re³¹, ñe³¹ zi³¹zi³¹ rə⁵³ lə³¹ ri³¹dɔ⁵³ da⁵⁵ dzɔ¹³² re³¹.
从前 猎人 一 有 COP 天 每 山 LOC 动物 打 去 COP
从前，有一个猎人，每天到深山里去打猎。

以上各分句标注 COP 之处，都不能用汉语“是”来对译。因此，藏语甘孜话文本中的系词都用 COP 来标注，让人看到系词在该语言中的显赫地位。当然，有些系词是否已经向语气词虚化，这是有待深入研究的课题。

（五）新的语法概念的引入

民族语言描写的深化，离不开描写框架的更新。面向人类语言多样性的类型学调查分析框架，是世界各地语言调查不约而同采用的框架，框架内部包含了很多西方和中国传统语法中没有或关注不多的语法概念，如现实性范畴、示证范畴等。丛书在采纳新标注法的同时，也面向不同语言的描写需求引入了不少新概念，实现了描写框架的一次更新换代，有益于准确认识相关成分的性质功能、深化民族语言的描写和跨语言比较。

前文实例中已涉及一些新概念的应用，如“定指标记、话题标记、示证范畴”等。丛书引入的新概念远未止于此，从各书前的缩写符号表，就可以一窥大概。下面再举一例——定语标记。汉语普通话中纯粹的定语标记就是虚词“的”，以往的标注常见以“的”来对译民族语言中的定语标记。但在世界语言中，定语标记常因定语性质的不同而异，例如英语中领格后缀's、前置词 of、关系代词 that 和 which 等都有定语标记的作用。很多语言定语标记还与名词化标记同形，汉语“的”既作定语标记，对应古汉语“之”，也兼名词化标记，对应古汉语“者”。因此，定语标记可以分解为不同的小类，如形容词后缀（如英语 national 之 -al）、领属（格）标记、关系从句标记、名词化标记等。这些细化的新概念，对于定语标记多样化的语言来说特别重要。藏语甘孜话就有多个定语标记，有时还可以连用，如果仅用“的”来对译，就会出现“的的”这样奇怪的标注。如（燕海雄、江荻 2016:80、72、76、67、68）：

- (15) dzɔ³¹pi⁵⁵ tshu⁵⁵ ze³¹ gə³¹ tei⁵⁵ tei⁵⁵ tshi⁵⁵ ly¹⁴.
土匪 抢 NOM GEN 财产 PL 回 找

祈求菩萨把土匪抢走的东西找回来。

(16) bu¹⁴ te³¹ tu³¹ she⁵³ tsha⁵³ khe⁵⁵ gə³¹ ra¹³² te³¹ tshu⁵³. 抢走了猎枪和打死的野羊。
枪 DEF COOC 杀 PFV NOM GEN 山羊 DEF 抢

(17) ɲa¹⁴ su⁵⁵ lə³¹ sa³¹ dze³¹ thu⁵⁵ dze³¹ ma³¹ ke⁵⁵. 我没吃没喝。
我 家 LOC 吃 NOM 喝 NOM NEG 有

(18) ri³¹ tɔ⁵⁵ ta⁵⁵ khe⁵⁵ tsı⁵³. 一个很穷的猎人
动物 打猎 NOM 一

(19) ho³¹ tei³¹ cha⁵⁵ gə³¹ za³¹ jo¹⁴ ɲu¹⁴ o³¹ gy¹³²? 你们买鹿角吗?
2PL 鹿 GEN 角 买 INT 要

对应例(15)“土匪抢走的东西”的“的”，甘孜话连用了两个定语标记，名词化标记 ze³¹ 和领属语标记 gə³¹（此处领属语已扩展到更广义的定语，包括名词化了的关系从句）；例(16)则连用了名词化标记 khe⁵⁵ 和领属语标记 gə³¹。例(17)-(19)是这几个标记单用的情况。dze³¹（疑为 ze³¹ 的变体）在例(17)中单独用作名词化标记，把动词“吃、喝”变成名词“吃的、喝的”；khe⁵⁵ 在例(18)中单独作名词化标记，把动词短语“狩猎动物”变成名词“猎人”；gə³¹ 在例(19)中单独作领属标记，“鹿之角”，体现其原型功能。可见，藏语甘孜话要让动词性成分充当定语，须连用两个定语标记：先用名词化标记将动词性成分名词化，再用领属标记作定语标记。如果采用对译法以两个“的”标注或单纯标为定语标记，就无法理解这两个标记的不同性质和功用，也显得怪异。正是名词化标记(NOM)、领属标记(GEN)这种新概念的引进，让文本标注提供了更精准的语法信息，提升了文本的语言学质量。

(六)逐词标注与整句翻译的张力

逐词标注与整句翻译有不同的要求。逐词逐语法素标注要尽量体现该成分的恒定属性，整句翻译则要顾及语境因素，使文本中的每一个句子都达到句子顺畅及语篇连贯，同时便于读者通过目标文本体会原文本中各成分包括语法素的表达功能。这样，就必然会造成标注行与翻译行的细节差异，形成两者的一种张力，这种张力是我们认识静态语言单位的动态功能的良好窗口。所以，标注的要求可以简要表述为“词有定类，依句辨用”。丛书总体符合这一原则，发挥了这一张力的效能。兹举土家语一例，只引原文正字法行，省去国际音标行（徐世璇等 2017:100-101）：

(20) ai⁵⁵le⁵³ go³⁵ jiu⁵⁵ a²¹ca²¹ ga²¹ ka²¹ ta⁵³ la⁵⁵, ka²¹mon²¹ yiu³⁵ ci⁵³,su⁵⁵ku⁵⁵ yiu³⁵ta³⁵xi⁵⁵,
TRAC 3sg 就 悬崖 上 柴 砍 ASPP 树 又 大 柴刀 又 钝
她在悬崖上砍柴，树又大，刀又钝，
cai²¹ mo⁵⁵ na³⁵xe⁵⁵ ta⁵³liau⁵³wa⁵⁵, ai⁵⁵do⁵⁵ta⁵³mo²¹ de²¹da⁵⁵, ai⁵⁵le⁵³xe⁵⁵qi⁵⁵ yiu³⁵tai³⁵,
好 STRP 一些 砍 ASPP INJ 还 砍 STRP 断 没 TRAC 力气 又 没有
砍了好一阵，还没砍断，力气也没有了，
ai⁵⁵le⁵³go³⁵jiu⁵⁵ka²¹mon²¹gu²¹lu⁵³,su⁵⁵ku⁵⁵je³⁵da⁵⁵ji²¹ka²¹bo⁵⁵mo²¹ka²¹mon²¹gu²¹lu⁵³.
TRAC 3sg 就 树 上 DIRP 柴刀 腋下 夹 ASPP STRP 树 上 DIRP
于是她就爬上了树，把刀夹在腋下往树上爬。

例(20)原文几处第三人称代词(3sg) go³⁵，译文中都根据实际所指译为“她”，但标注并不适合用“她”，因为该词本身并没有性别义。再分行看。第一行，TRAC 是承接连词，而汉语意合复句盛行，并不需要都译出。“悬崖上”原文类似后置词结构，译文加上前置词“在”

更加自然，因为汉语是前后置词并存配合的语言。动词 ta^{53} “砍”原有体标记 la^{55} ，表示完成体，译文没有带体标记“了”，因为在表达连续行为时，汉语并非每个已然行为动词都要加“了”，一般只在凸显动作结束或现实相关性时采用。第二行两次使用结构助词（STRP） mo^{55} ，这是汉语所不需要的，特别是动词和结果补语之间在普通话中不能用补语标记，因此译句中都空缺。第三行，“夹柴刀”和“上树”两个行为之间，土家语要用结构助词连接，可见连动式不如汉语发达，而汉语中这样的连动式很自然，不能插入连接性虚词。上树的动作，土家语用路径动词 gu^{21} “上”加趋向词 lu^{53} 表示，体现了位移事件的“动词框架”类型；译文采用“往树上爬”，由方式动词“爬”作谓语核心，体现位移动词更强的“卫星框架”类型。尽管土家语是一种受汉语影响较大的藏缅语，但是从一小段文本的标注和译文的差异，还是可以看出二者诸多类型差异。这也是分行标注法的学术优势所在。

二 丛书文本标注法的尚待改进之处

作为国内第一套采用新型分行标注法的文本丛书，丛书为中文著作的民族语言语法例句和文本标注探索出了一条新路。因为是首次尝试，且参与者众多，所以不一致、不完备之处也在所难免。下面我们将一些有代表性的问题进行分类例析，提出个人意见，供学界参考。

（一）分词连写不统一，切分不充分，逐词逐语法素标注未充分实现

新型分行标注法的核心内容是逐词逐语法素标注，实词切分注释到词，虚词注明功能，形态切分到构形语法素并进行功能说明。

为此目标，拼音或音标首先应按词切分和连写，即词间分开，词内连写，这样也方便读者知晓下一行的释义对应哪个词。目前，丛书在这一点上尚未统一。有些卷册按此标写，有些卷册则以音节为单位拼写，而释义则仍按词进行，导致词义和词形对应不严密。独龙语（杨将领 2020）、黎语白沙话（吴艳 2021）、土家语（徐世璇等 2017）都是按词切分连写的。再如壮语（蓝利国 2016）、嘉戎语卓克基话（林幼菁 2016）等也均按词连写。但是，藏语的几册文本都没有按词切分连写，总体都以音节为拼写单位，但具体做法有别。藏语拉萨话（江荻 2016）完全按音节拼写，词内音节之间、词与词之间，间隔相同，都是半角码距离。藏语甘孜话（燕海雄、江荻 2016）和藏语噶尔话（龙从军 2021）也按音节拼写，但是大致采用了两种间隔，词内音节之间是半角码距离，词与词之间是全角码距离，但是甘孜话的词界距离并不稳定一致。先看拉萨话一例，取原文行，分析行略去（江荻 2016:76）：

(21) phyug po der bong bu rkyang dkar zhig yod pa dang. dbul phongs der wa la zhig yod.
富 那 驴 浅棕色 一 有 NMZ 和 穷人 那 青灰色 一 有
那个富人家有一个浅棕色的驴，那个穷人家有一个青灰色的驴。

上例双音节词 $phyug$ po “富”（指富人）、 $bong$ bu “驴”、 $rkyang$ dkar “浅棕色”、 $dbul$ phongs “穷人”、 wa la “青灰色”，词内两个音节间的距离跟其他词的词间距离相等，不易一眼辨认出哪些音节是一个词。词的释义都对齐首音节，但是从下一个词的释义位置可以看出其前的音节仍属于上一个词，如 $rkyang$ dkar “浅棕色”前的音节 bu 仍当属于“驴”一词。

藏语甘孜话和噶尔话似乎用空格距离区分词内音节间隔和词与词的间隔，但前者做得并不严格，因此据此无法全部划出词的界限。为了直观地观察词界划分的处理方式，我们截图选择甘孜话一例（燕海雄、江荻 2016:86）：

(22)

tsha⁵⁵ te⁵⁵ ji¹⁴ ge³¹ ja³¹ de³¹ tu⁵⁵ tsha⁵⁵ te⁵⁵ ji³¹ ge³¹ te⁵⁵ de¹⁴,
税 告示 贴 去 税 告示 上面

去张贴征税的告示。告示上写着：

tchā⁵⁵ tshu⁵⁵ re³¹ re¹⁴ lō⁵⁵ n̄e⁵⁵ re¹⁴ tse?⁵³ gy¹³² lō⁵⁵ n̄a⁵⁵ ma³¹
家 每 猫牛 每 交 要 猫牛 H-NEG

nə³¹ ta⁵⁵ yə³¹ tsha⁵⁵ ce⁵⁵ gy¹³².

如果 马 INSTR 代替 要

“每户人家要缴猫牛一头，无牛者得以马代替。”

上例中，tsha⁵⁵te⁵⁵ “税”的两个音节间只有不足半角字符的距离，ji¹⁴ge³¹ “告示”的两个音节间也是如此，而tsha⁵⁵te⁵⁵ 和 ji¹⁴ge³¹ 之间则有1个全角符的距离，且“告示”的释义从te⁵⁵结束以后的下方开始，这样可以推出两个词的词界。其他词语大致也是如此，但是并不尽然。如第二行 tshu⁵⁵ 和前面 tchā⁵⁵ “家”距离不足一个全角字，与后面 re³¹re¹⁴ “每”则只有一个半角字，难以看出该音节属前还是属后。“每”的释义是从 re³¹ 音节下开始的，其前 tchā⁵⁵ 本身能独立表示“家”义，而后文“每”只用 re¹⁴ 表示。因此，读者难以判断 tshu⁵⁵ 应归哪个词。再看第一行 ji¹⁴ge³¹ “告示”和 ja³¹de³¹ “贴”明确分属两个不同的词，但是距离只有一个半角符。可见，按间隔距离划分词界在此书中未贯彻到底。相较而言，藏语噶尔话文本（龙从军 2021）贯彻距离定词界的原则比较彻底，可能是后出转精。不再举例。

应该看到，广义的汉藏语系语言包括藏语的不同方言，都是程度不同的音节显赫语言（刘丹青 2018；孙克敏、刘丹青 2020），音节界也有重要的语言学意义。由于藏语存在复辅音的声母和韵尾，因此按词连写会导致某些音节界限的模糊，藏文传统也是按音节拼写的。这可能是藏语几个方言都采用按音节切分方法的原因之一。但需注意，藏语拉萨话是按无声调语言标注的，甘孜话和噶尔话是按声调语言标注的。一标注调号，音节界就绝对清楚了。因此，至少藏语甘孜话和噶尔话应当按词连写，以同丛书体例一致。至于拉萨话，如要保留音节界，则可以采用统一的半角距和全角距的对立来区分是词内音节间停顿还是词间距离。

按词切分连写是总原则，在遵循该原则的前提下，有些卷册在操作层面仍有改进余地。

土家语是按词切分的，但是切分出的个别词或形态要素没有标注。下面两例只引正字法行，省去音标分析行（徐世璇等 2017:96、91）：

(23) ai⁵⁵ biu³⁵ de⁵⁵ hai⁵⁵ khuei³⁵ ya⁵⁵. 那女孩这么小。

那 女儿 EMP 还 小

(24) ai⁵⁵le⁵³ nga³⁵ ne⁵⁵ a²¹ne⁵³ qin⁵³mo²¹ gao⁵³-i³⁵ yo⁵⁵. 我母亲怎么办呐。

TRAC 1sg STRP 母亲 怎么 搞 INJ

例(23)句末的 ya⁵⁵[ja⁵⁵]没有标注。我们从其他例句中看到，ya⁵⁵ 标为 INJ，即感叹语气词。作者应该是漏标了。例(24) gao⁵³-i³⁵ 标注为“搞”，这只是该动词（当为汉语借词）的词义，后面用短横连接的 -i³⁵ 是书中多次出现的疑似形态成分，但均未作标注，亦属漏标。

土家语卷还存在因译文部分脱落重要成分造成意义误解的，如（徐世璇等 2017:113）：

(25) goe⁵³ an⁵⁵ngqi⁵³ loe⁵⁵, na⁵⁵ poe⁵⁵ na⁵⁵ poe⁵⁵ ai⁵⁵ la³⁵tu⁵⁵ ka⁵³ bo⁵⁵.

3pl 弟弟 TOP 一 夜 一 夜 那 南瓜 看守 ASPP

整晚守着那些南瓜。

例(25)句中用话题标记引出的话题“他弟弟”没有在译文中出现。问题是，此句是一个话题转换句，它的上文译文是“而他的哥哥呢，没娶媳妇，整天在家玩，米嘛，吃完就打（用打米机），吃完又打，整天睡在家什么也不做，于是粮食也不多了”。本句开始话题转为“弟弟”，由于本句话题被略去，从译文看，仍在承袭上文叙述“哥哥”，这是错误的。

藏语甘孜话也存在标注缺漏的情况。如燕海雄、江荻(2016)第111页和116页都有句末系词 re^{31} 后的一个成分 su^{53} 没有标注，而 re^{31} 和 su^{53} 间的距离都是全角符的词间距离。

(二) 标注不统一，同一成分同一用法标注不一

标注不统一，主要发生在多义多功能虚词上。这类词的义项划分本身就很棘手(详见下文)。此处要指出的是，一些词项在同样的意义功能上却被标注为不同的属性，这是需要避免的。至少在同一部著作中，同类的意义功能应统一标注。发现这类问题有赖于对标注材料的详尽深究，我们重点考察的卷册有限，这里主要以藏语甘孜话为例。

ni^{31} 是甘孜话一个高频出现的连接谓词性成分的连词，可见甘孜话连动式不显赫，谓词连接要靠标记。该标记仅在燕海雄、江荻(2016:62)就出现了4次，其中3次都被标注为SEQ(顺时连词)，1次被标为ADV。缩写ADV在略语表中漏注，我们理解为状语(adverbial)标记、状语助词。试比较：

(26) $a^{55}\text{pha}^{55}\text{te}^{31}\text{she}^{55}\text{du}^{31}\text{ni}^{31}\text{təhi}^{55}\text{nū}^{53}\text{lu}^{14}$. 老爹爹带着极大的愁苦回到了家里。

爸爸 DEF 心 悲伤 ADV 家 里 回

(27) $\text{su}^{55}\text{ni}^{55}\text{ŋa}^{14}\text{dza}^{33}\text{bo}^{53}\text{gsha}^{55}\text{dzo}^{132}\text{ni}^{31}\text{kho}^{53}\text{lə}^{31}\text{ne}^{31}\text{dza}^{33}$.

明天 我 国王 地方 去 SEQ 3SG ALL 回答

明天我到国王面前去回答他。

汉语译文两句均为连动句，但可能作者认为“带着极大的愁苦”带有修饰“回家”的功能，语义上更像状语，于是改用ADV来标注。实际上，连动式内部各项和并列结构内部各项语义关系略有轻重很正常，如汉语的连动式(“笑着说”)和英语的并列式(bread and butter“涂黄油的面包”；wait and see“等着瞧”)。例(26)是顺时连词的正常用法之一，不需要另外标注为状语标记，给该虚词增加复杂性。况且全书大量 ni^{31} 的用例，除了此例，全部都标注为SEQ，其中也包含了比例(26)更像状中结构的例子，如(燕海雄、江荻2016:131)：

(28) $\text{lə}^{14}\text{nu}^{14}\text{lə}^{31}\text{kha}^{55}\text{tei}^{31}\text{tee}^{55}\text{je}^{31}\text{me}^{31}\text{ma}^{31}\text{ta}^{53}\text{ni}^{31}\text{dze}^{31}\text{po}^{55}\text{ee}^{55}\text{ni}^{31}\text{sh}^{55}$.

路上 LOC 3PL 才 贫富 NEG 看 SEQ 和谐 SEQ 走

路上他们不分贵贱，非常和好地走着。

此例用了两次 ni^{31} ，汉语译文直接用状语助词“地”来标注这两个语义上似乎修饰“走”的成分，其状语性强于例(26)中用ADV标注的 ni^{31} 前的成分，但是标注仍是SEQ。这是合理的，应当统一如此，例(26)标为ADV影响了标注的内部一致性。

再来看施事标记 yə^{31} 。施事标记是很多藏缅语中都存在的语法要素(往往同时兼工具标记)。 yə^{31} 在甘孜话里也很常用，但是像在其他藏缅语里一样，它远非强制性的，是一种施事的强化标记，在语用需要时使用。在甘孜话文本里，它偶尔还被标记为领属标记GEN^①。如

^① 燕海雄、江荻(2016)标注为GEN的，实际上用于更广义的定语，不限于领属语，见上文例中的 $gə^{31}$ 。至于发音很近的 yə^{31} 和 $gə^{31}$ 是否同源，存疑。

(燕海雄、江荻 2016:148):

- (29) tʂə⁵⁵ly⁵³ yə³¹ dʐə³¹ te³¹ ti⁵⁵ju³¹ki⁵⁵ ne⁵⁵. 笛子的声音更加动听。
笛子 GEN 声音 DEF 更加 动听

如果 yə³¹ 确实有领属标记或定语标记的用法, 那例(30)中后一个 yə³¹ 更应该是定语标记, 而不宜标为施事标记(燕海雄、江荻 2016:149):

- (30) kɔ⁵⁵teɔ⁵⁵ yə³¹ tsj⁵³ ta⁵³ dʐi⁵⁵ tʂə⁵⁵ly⁵³ be³¹ khe⁵⁵ yə³¹ pe⁵⁵ te⁵⁵ tu³¹ dʐə³¹tsa⁵⁵re¹⁴ tɔ³¹.
公主 AGT 一 看 以后 笛子 吹 NOM AGT 儿子 DEF COMT 像 COP SF
公主一见, 正和吹笛子的那个青年一模一样。

从前文例(15)(16)的分析可知, 甘孜话关系从句一般需要连着带名词化标记和领属标记。例(30)中后一个 yə³¹ 正用在关系从句(“吹笛子的”)和名词化标记之后, 跟例(15)(16)中的领属标记 gə³¹ 相同。因此, 这个 yə³¹ 当与例(29)中的 yə³¹ 一样, 标注为领属语标记 GEN, 况且它在此处位置绝对不应标注为施事。

最后说一下甘孜话中功能最多样 lə³¹。此语法成分在文本中分别被标注为 LOC(位格)、ABL(从格)、ALL(向格)、DAT(与格)、OBJ(对象格)、话题标记(TOP)6种属性, 后文还会就此讨论。这里只举一对功能同而标注不一之例(燕海雄、江荻 2016:49、52):

- (31) eo⁵⁵a⁵⁵ tɕi³¹ lɔ⁵⁵pɔ⁵⁵tɕe⁵³ gə³¹ na⁵³ nū⁵³ lə³¹ dʐə¹⁴bɔ¹⁴ dʐu³¹ la⁵⁵ re³¹.
老鼠 PL 大象 GEN 鼻子 里 ALL 钻 进 IPFV COP
小老鼠钻进大象的鼻子里去了。
(32) eo⁵⁵kha⁵³ dʐo¹³² jo¹⁴ tu³¹ ee⁵³ la⁵⁵ re³¹ ti⁵⁵ma⁵⁵hu⁵⁵ dy¹²³ tsj⁵⁵ nū⁵³ lə³¹
狼 走 CONT COOC 说 IPFV COP 忽然 洞 一 里 LOC
lɔ⁵³ dʐə³¹ re³¹. 狼一边走着一边说, 忽然掉到一个山洞里去了。
掉 进 COP

两句都表示动物到一个空间的内部, 但是, 同样表终点的标记 lə³¹, 在例(31)中标为 ALL(向格), 在(32)中标为 LOC(位格), 当同而异了。

(三) 上位代替下位, 语法性质定位过于笼统

按照逐词逐语法素的标注要求, 不但实词要标注词义, 虚词和形态成分也要标注其语义和功能, 正如名词不能只标出名词, 而要标出词义, 语法素也是如此, 只是一般不便逐词直译, 而是对所属语法语义范畴进行描写。丛书大致循此实践, 但实现程度不等。范畴有个层次问题, 作为文本标注, 应当注出该语法素特有的范畴义, 只有真正同义的语法素才可以共享标注的范畴。嘉戎语卓克基话(林幼菁 2016)的标注就特别细致, 很多语法素都得到了精确刻画。当然, 在掌握语义功能的异同上, 各书宽严可以略有差异。如藏语甘孜话多功能标记 lə³¹的功能之一是 ABL(从格), 见例(31); 文本中另有一个虚词 n̩³¹, 也标注为 ABL(从格), 见燕海雄、江荻(2016:146)。这两个标记的语义、功能或使用条件也许不尽相同, 不一定能无条件互换, 但如果都属于从格通常表达的语义, 则两者都标为 ABL 是可以的。

但是, 有些语法素标注了过于上层的范畴, 实际覆盖了很多不同的甚至互相对立的语义功能语法素, 如一些宽泛的词类名称或大范畴的名称。这样的标注就不够到位, 信息较粗疏, 影响语料质量。下面以土家语为例。徐世璇等(2017)略语表收录的标注代码为23个, 不到嘉戎语卓克基话卷的一半, 后者是50个。其中的重要原因是嘉戎语综合性强, 形态范畴更多、更细, 但前者提取的范畴太宽泛也是一个原因。该书过于依赖词类名称和宽泛的范畴名称。

如“结构助词、体助词、比较助词、趋向助词、强调助词、格标记”，其中的体助词包含了多个相互对立的体，例如完成、持续、进行等，不宜用同一个标签来标注。再如格助词，里面包含相互对立的格，却用同一个缩写代码 CASP 标注，读者无从得知带格标记的名词的句法地位和该语言的格范畴系统。现据我们对语料标注的初步观察，将各个格助词的功能分析如下，仍录正字法行，略去音标行（徐世璇等 2017:116、73、54、46、174）：

- (33) jiu⁵⁵ goe⁵³ go³⁵ ai⁵⁵ la³⁵ tu⁵⁵ hu²¹ pa²¹ tu³⁵ bo⁵⁵ po⁵⁵ lu⁵³.
就 3pl CASP 那 南瓜 河 里 CASP 滚 DIRP
然后他们就把那南瓜滚到了河里。
- (34) a. no⁵³ ci⁵⁵ ba⁵⁵ ne⁵⁵ hhii⁵⁵ da⁵⁵ ye²¹ na³⁵ be⁵⁵ wo⁵³ le⁵³ cu⁵⁵ bo⁵⁵ son⁵⁵ ko⁵⁵ diu⁵⁵.
大人 两个 一起东西一些 背 ASPP 家 CASP 回 DIRP
大人俩背着一些东西回来了。
b. jin⁵⁵ can²¹ go³⁵ bo⁵⁵ jin³⁵ xian⁵⁵ a⁵⁵. (经常) 给她烧香。
经常 3sg CASP 进香 INJ (引者：原译文无“经常”)
- (35) a. hu²¹ de²¹ ko⁵⁵ ba⁵⁵ bi³⁵ cen⁵⁵ le⁵³ go³⁵ go³⁵ ga³⁵-i³⁵. 头一伸要咬她。
呼 STRP 头 小 伸 ASPP 3sg CASP 咬
b. ai⁵⁵ le⁵³ ri⁵⁵ ma⁵³ xe³⁵, la⁵⁵ sa⁵⁵ de⁵³ go³⁵ nga³⁵ ne⁵⁵ ri⁵⁵ bo⁵⁵ la⁵⁵ man⁵⁵.
TRAC 做的人 有 有的人 CASP 1sg STRP 做 ASPP ASPP IND
有人种，有的人种了我的土豆。（引者：“土豆”是语境中的信息）

上例是一句之中同时用 go³⁵[ko³⁵]、bo⁵⁵[po⁵⁵]和分别用这两者的。从例 (33) (34a) 可见，bo⁵⁵[po⁵⁵]应该是向格（终点格），分别是滚的方向——到河里，回的方向——到家。向格也往往扩展到接受者（如英语 to，日语に ni），所以也用于例 (34b)，表示进香的接受者——灵姑神。文本中大量的 bo⁵⁵ 用例大多是用于向格的。go³⁵ 在 (33) (35b) 中都用于施事，但 (35a) 中 go³⁵ 却似乎用于受事（咬的对象），但如果其前的第三人称代词 go³⁵ 指蛇，则格标记 go³⁵ 仍可暂定为施事强化标记。此标记性质尚待进一步确定。

例 (36) 是 do²¹[to²¹]用例（徐世璇等 2017:75）。由于用例太少，我们暂时推测为引介被动句施事的格。

- (36) zi⁵³ ji²¹ moe⁵⁵ soe³⁵ ta⁵⁵ a⁵⁵ ma⁵⁵ do²¹ ga³⁵ lu⁵³. 猪腿被邻居奶奶吃掉了。
猪腿 TOP 旁边的 奶奶 CASP 咬 DIRP

例 (37) 是 da⁵³[ta⁵³]的用例（徐世璇等 2017:74、70、58）：

- (37) a. ai⁵⁵ le⁵³ goe⁵³ jiu⁵⁵ goe⁵³ ne⁵⁵ bo⁵⁵ li²¹ da⁵³ si⁵⁵ le²¹ la⁵⁵. 于是他们就问孩子。
TRAC 3pl 就 3pl STRP 孩子 CASP 问 ASPP
b. ai⁵⁵ le⁵³ ai⁵⁵ na⁵⁵ poe⁵⁵, go³⁵ cu⁵⁵ a²¹ ne⁵³ da⁵³ li²¹ la⁵⁵.
TRAP 那 一 夜 3sg 家 母亲 CASP 说 ASPP
有一个晚上，他回到家跟妈妈说。
c. go³⁵ bu²¹ ci³⁵ hu⁵⁵ so⁵³ da⁵³ da⁵³ go³⁵ liao⁵³.
3sg 不次乎（地名）人 CASP 吵架 ASPP
他跟不次乎（地名）人吵架了。

da⁵³ 的语义功能基本上符合伴随格（comitative）或指人对象格的属性，用于人际互动性行为，主要是言语行为。例 (37a) (37b) 是一方对另一方的言语行为（问、说），更像对象

格; (37c) 是相互行为(吵架),更像伴随格。伴随格兼表单向行为或对象格用于互动性行为,都比较常见,所以定性为这两者中的任何一种都可以。

下面是双音节格标记 $ho^{21}le^{21}$ [$ho^{21}le^{21}$] 的例子(徐世璇等 2017:106、143、65)。其属性比较明显,就是工具格:

- (38) a. $goe^{53} biu^{35} jiu^5 su^{55}ku^{55} ho^{21}le^{21} go^{35} ga^{53} la^{55}$. 他们女儿就拿刀砍她。
3spl 女儿 就 柴刀 CASP 3sg 砍 ASPP
b. $san^{55}mi^{55} qe^{55}hhoe^{55} ho^{21}le^{21} ri^{55} do^{55}$. 团馓要用糯米做。
团馓 糯米 CASP 做 要
c. $han^{35}ri^{55}, an^{35} ho^{21}le^{21} no^{53} qian^{35} do^{55} suan^{35} ta^{53} ya^{55}$.
这样 做 1pl CASP 人 一点 都 算 不 INJ
他这么对我们,完全不把我们当人。

最后是 le^{55} ,用于空间处所和时间,应该是处所格 LOC。用例不多,因为本身带空间义的名词在土家语文本中常常直接充当处所状语,其中有些方位成分可能已经有后置词的性质了,但是作者仍按方位名词标注。例如(徐世璇等 2017:66、139):

- (39) $moe^{21}kii^{21}coe^{55} wo^{35}tu^{55} le^{55} zao^{35}ne^{55}$. 打心底可怜(他)。
肚子 里面 CASP 可怜
(40) $lau^{35}tsi^{55} mi^{35}ne^{53} le^{55} ga^{55}gon^{55} ka^{21}bu^{21} ba^{53} di^{55} hu^{53}$.
明天 后天 CASP 外公 外婆 看 DIRP ASPP
过两天外公外婆要来看月了。

上例标注为 CASP 的成分,按标注惯例和丛书的常规,宜用具体格的名称缩写来标注,如: AGT(施事格)、LOC(位格)、ABL(从格/源点格)、ALL(向格/终点格)、INSTR(工具格)、COMT(伴随格)、PAGT(passive AGT: 被动施事格)。

前文提到的土家语其他一些高层范畴标记,如结构助词、体助词、趋向助词等,也应当像上述不同的格标记一样,标注出范畴内的功能义。例如,体范畴标注应当像藏语甘孜话那样用完成体、持续体等具体概念的缩写,而不是笼统的 ASPP(体助词);再如趋向助词,也可以像嘉戎语卓克基话的趋向词缀一样标出每种趋向的方向:朝上、朝下、朝上游、朝下游、朝东、朝西。上文之所以选格标记为例来分析,一是因为格标记(有些也可以理解为介词性的后置词)的功能特别重要,是句子基本句法构造的重要外显表征;二是因为徐世璇等(2017)第一章“语法导论”部分对名词的格完全没有提及,读者在文本中看到 CASP 时,无法在导论部分查到相关知识点;而有些范畴在“语法导论”中已有所介绍,读者可以对照阅读,如该书第 20-23 页各种体、貌的详细描写,第 27 页对趋向助词的具体介绍。

(四) 概念错位,语法性质定位存在偏差

丛书的大部分语法素标注都是值得肯定的;上面所说的问题,主要是单位大小和标注范畴的粗细问题。不过,丛书语料标注有一个难点,即作者需要在选录的大量文本中对每一个单位都作出定性和标注,无法回避。这事实上设定了对该语言研究水准的较高要求。因此,很难完全排除标注定性不当的情况。这种情况尤其容易出现在多义语法素上,因为它们使用频率高,用法多样,义项划分难度大。如果义项分得粗,将多种功能视为同一义项的语境变体,则是保险处理,不易有大错,但对语法现象的刻画难免粗疏;如果分得细,可以更加全面深刻地揭示语法素的功能,但是,很难做到处处定性精准。

下面以藏语甘孜话的多功能标记 *lə³¹* 为例。*lə³¹* 在文本中分别被标注为 LOC(位格)、ABL(从格)、ALL(向格)、DAT(与格)、OBJ(对象格)、TOP(话题标记)6种属性。对照文本语料，我们认为前5种功能的设立是合理的。当然，在颗粒度上有简化的余地：位格、从格和向格这三种可以合并为位格 LOC。在语言库藏并无三种格的形式区分时，广义的空间处所标记兼表源点和终点是常事，如苏州话就能用一个前置词“勒”[lə²³]表示处所（“勒食堂里吃饭”）、来源（“勒抽屉里拿出一本书”）、方向终点（“甩勒河浜里”），它们都是 LOC 在语境中的具体实现，确切解读是配合句内成分尤其是动词的事件结构来完成的。事实上甘孜话目前的处所标记三分标注得也并不严格，上引例(31)(32)，*lə³¹* 同表事件终点，一个标 ALL，一个标 LOC，说明两者本来相通，不能说哪种标注错了，但两者没有理由不一致。当然，细分为三尚在合理范围，尤其在语义地图模型中，从细区分是常态，也方便解释从方向终点到与格和对象的路径。但是，我们对这个语法成分的话题标记功能有怀疑。虽然在话题优先的汉藏语言中，话题标记来源可以是多样的，但是，话题跟动作对象共享标记，在语法化路径中是罕见的。通过对语料的分析，可以质疑这一定性。

lə³¹ 被标为 TOP 的句子有两个突出表现：①出现频率很低，全书只发现 5 例，远少于该语法成分其他各功能的用例；②适用的谓词语义高度受限，基本上局限于含“拥有”义素的动词。例如（燕海雄、江荻 2016:48、69、97、76、127）

- | | |
|---|-------------------|
| (41) eo ⁵⁵ a ⁵⁵ lə ³¹ tha ⁵³ ma ³¹ re ³¹ . | 老鼠没有办法。 |
| 老鼠 TOP 办法 H-NEG COP | |
| (42) ɳa ¹⁴ lə ³¹ tsa ⁵⁵ ta ⁵⁵ wi ¹⁴ ja ⁵⁵ təha ⁵⁵ mə ³³ pə ⁵³ ma ³¹ . | 我没有像头发那么多的钱。 |
| 1SG TOP 头发 像 NOM 钱 多 H-NEG | |
| (43) nə ³³ ma ⁵³ dza ³³ bə ⁵³ lə ³¹ kə ⁵⁵ teo ⁵⁵ tei ⁵³ jə ¹⁴ a ⁵⁵ re ³¹ . | 太阳国王有一个公主。 |
| 太阳 国王 TOP 公主 一 有 IPFV COP | |
| (44) ɳa ¹⁴ su ⁵⁵ pa ⁵⁵ lə ³¹ lə ⁵⁵ re ³¹ lə ³¹ be ³¹ le ⁵⁵ ni ¹⁴ re ³¹ . | 我家里的一头牛，一年能生两只小牛。 |
| 1SG 家 牛 TOP 年 每 ABL 小牛 两 COP | |
| (45) təhe ⁵³ lə ³¹ dza ³¹ lu ⁵⁵ tsa ³¹ la ⁵⁵ re ³¹ . | 大哥叫甲鲁。 |
| 大 TOP 甲鲁 叫 IPFV COP | |

由上可见，这个话题标记在全书中只出现了 5 次，其中 4 次都是以“有无”类语义为谓语。其中例(44)的“生小牛”其实并没出现现实义动词，只用了系词，系词也有存在义，这里也就表示有了小牛。话题结构是一种语义上比主谓结构更加松散的结构，对语义的限制很少。而“有无”类是谓语众多语义类别中的一个独特小类。话题标记的使用不会如此狭窄。我们需要关注该标记与有无义之间高度关联的原因。其实原因并不复杂。*lə³¹* 由介引空间处所、尤其是方向终点功能发展出了与格的功能，这是人类语言极其常见的语义扩展路径，英语的前置词 to 和日语的后置词兼与格标记“に”ni，都兼介引处所方向和与事的功能。燕海雄、江荻（2016）列出的 *lə³¹* 的功能之一就是与格。在很多有与格的语言中，领有句的领有主体是由与格而非主格充当的，如德语（鲁道夫·霍贝格、乌尔苏拉·霍贝格 2014:368）：

- | | |
|-------------------------------------|----------------------|
| (46) Ihm gehörten große Ländereien. | 他拥有大片田地。 |
| 他.与格 拥有 大的 田地 | (引者：标注方法由本文根据德语情况调整) |

与事的核心语义是接受者/得到者，得到即开始领有，领有即已经得到，两者语义相通，

因此在与格语言的认知图式中，领有者和得到者被赋予同一格位。与格标记 $lə^{31}$ 用作领有句主体论元标记非常自然，符合与格常态，不必另外离析出一个话题标记功能。事实上该书中有关类谓语句中的拥有主体论元，也有被标上 OBJ(对象格)的。例如(燕海雄、江荻 2016:56):

- (47) $\eta a^{14} lə^{31} \text{ ehy}^{55} təhə^{55} pə^{55} jə^{31} rə^{31}$.
1SG OBJ 力气 大 有 COP

此句的句法结构与例 (41-43) 的领有句完全一致，标注领有主体的 $lə^{31}$ 也应该统一标为与格 DAT，对象和话题标记的定性缺乏依据。

上面还有例 (45) 尚未解释，其实也有同样例证。一个人叫什么名字，就是拥有或被赋予什么名字。如汉语的双宾语“叫他老王”，名字的主人正占据间接宾语与事之位。同样，在与格语言里，“某人叫什么”，“某人”也取与格。如塞尔维亚—克罗地亚语(黄雨辰 1987:141):

- (48) Kako ti je ime?
什么 你.与格 是.3sg 名字
你叫什么名字?

注意此句中的 ti “你”是第三格(即与格，主格为 $tı$)，句法上不是主语，句子属非人称句，系动词 je 是第三人称单数形式，而非第二人称。所以例 (45) 中的“大哥”带与格标记在格语言中是很自然的。假如 $lə^{31}$ 真是话题标记，其搭配的谓语类别不会如此受限。

与这个标记密切相关的有一个 la^{31} ，与 $lə^{31}$ 仅元音微别。 la^{31} 在燕海雄、江荻 (2016) 中被标注为 TOP 的次数比 $lə^{31}$ 更多，有 11 次，另有第 200-201 页两例同类用法漏标。此外， la^{31} 也有标为 ABL(从格)、DAT(与格)的，其中实际上也有与标为话题标记的情况相同的。总体上， la^{31} 的各种功能都是 $lə^{31}$ 所具有的，而且其中标为 TOP 的，所搭配的谓语绝大部分与有无类语义有关。因此，可以推测 la^{31} 是 $lə^{31}$ 的一个语音变体，总体上就是处所—与格的标记。话题标记的定性未必符合其真实属性。限于篇幅，此不赘言。

三 余 言

本文主要是举例评论，小结可免，但可以再强调一下以下要点。

1. 标注文本须落实到每个实词和语法素(包括虚词和构形形态要素)。文本应分词连写，词内的不同语法素在标注时用短横隔开，同一语法素的不同范畴义用墨点分开。
2. 虚词词类或语法范畴下不同的功能成分，要分别具体标注，不能只标词类名称或范畴名称。主观性很强的语气词等，在尚未得到精确描写前，可以暂标词类名。
3. 多功能语法素可以适当划分义项，按义项标注。若一个义项可以概括，具体语义主要依赖语境信息的功能，也不必过度划分义项，可以用更加概括的功能来标注。
4. 同一语族、语支、语组的语言，尽量采用相近的标注模式。
5. 对人类语言多样性和类型学成果的了解多多益善。

参考文献

- 陈玉洁、Hilario de Sousa、王 健等. 2014.《莱比锡标注系统及其在汉语语法研究中的应用》，《方言》第 1 期.
- 黄成龙. 2005.《语法描写框架及术语的标记》，《民族语文》第 3 期.
- 黄雨辰主编. 1987.《塞尔维亚—克罗地亚语语法》，北京：外语教学与研究出版社.

- 江 荻. 2016.《藏语拉萨话语法标注文本》，北京：社会科学文献出版社.
- 蓝利国. 2016.《壮语语法标注文本》，北京：社会科学文献出版社.
- 林幼菁. 2016.《嘉戎语卓克基话语法标注文本》，北京：社会科学文献出版社.
- 刘丹青. 2018.《制约话题结构的诸参数——谓语类型、判断类型及指称和角色》，《当代语言学》第1期.
- 龙从军. 2021.《藏语噶尔话语法标注文本》，北京：社会科学文献出版社.
- 鲁道夫·霍贝格、乌尔苏拉·霍贝格编著. 2014.《杜登德语语法》，朱建华等编译，上海：上海译文出版社.
- 罗天华. 2023.《莱曼标注和语法标注的规范问题》，《当代语言学》第4期.
- 孙宏开、胡增益、黄 行. 2007.《中国的语言》，北京：商务印书馆.
- 孙克敏、刘丹青. 2020.《藏语音节凸显的库藏类型学考察》，《民族语文》第2期.
- 吴 艳. 2021.《黎语白沙话语法标注文本》，北京：社会科学文献出版社.
- 徐世璇、周纯禄、鲁美艳. 2017.《土家语语法标注文本》，北京：社会科学文献出版社.
- 燕海雄、江 荻. 2016.《藏语甘孜话语法标注文本》，北京：社会科学文献出版社.
- 杨将领. 2020.《独龙语语法标注文本》，北京：社会科学文献出版社.
- 中央民族学院少数民族语言研究所第五研究室. 1983.《壮侗语族语言文学资料集》，成都：四川民族出版社.
- Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison.
In Michael Tomasello (ed.), *The New Psychology of Language*, Vol. 2: 211-243. Mahwah, NJ: Lawrence Erlbaum.

A Historic Leap and Room for Improvement in Text Glossing: The Case of the Book Series of Grammatical Glossing of Texts in Minority Languages in China

LIU Danqing

[Abstract] The present paper notes that in the documentation and description of minority languages in China, grammatical glossing has been relatively weak. The book series of *Zhongguo Shaoshu Minzu Yuyan Yufa Biaozhu Congshu* (Grammatical Glossing of Texts in Minority Languages in China), edited by Jiang Di et al., however, makes a historic leap in this field, in adopting interlinearized glossing method based on typological view of language diversity, providing more sufficient and accurate syntactic and semantic information regarding the sentence structure, remarkably raising the linguistic value of the texts. The paper, with fine analyses of many examples from the book series, illustrates six advantages in this new practice of text glossing and points out four respects which await improvement. It concludes with some suggestions for grammatical glossing of texts.

[Keywords] minority languages text glossing word-by-word glossing grams

(通信地址: 518060 深圳 深圳大学人文学院)

【本文责编 吴雅萍】