

中文文本标注规范 (5.0 版)

**Tokenization Guidelines of
Chinese Text (V5.0)**

黄昌宁 李玉梅 朱晓丹

Chang-Ning Huang, Yumei Li, and Xiaodan Zhu

微软亚洲研究院

Microsoft Research Asia

2006年3月27日

March 27, 2006

第一章 概述

1.1 版本说明

微软亚洲研究院《命名实体标注规范》3.0 版是为 30 万词《人民日报》语料的命名实体（NE）标注任务制定的。其英文版‘Guideline on Chinese Named Entity Annotation’成稿于 2003 年 2 月，用于 LSP（Lexical Service Platform）课题。当时在研究院，命名实体识别（Name Entity Recognition）和自动分词（Word Segmentation）是文本处理中互相独立的两个过程，所以未曾深入考虑分词词表（lexicon）对命名实体标注带来的影响。2005 年 3 月至 7 月在准备第二届国际自动分词评测（SIGHAN Bakeoff 2005）的 237 万词训练语料的过程中修订了该规范，形成 4.0 版。《命名实体标注规范》4.0 版的一个最大特点是把命名实体识别有机地融入到中文自动分词的整体过程中去。因此，除了命名实体自身的定义以外，还需要系统地阐明词表词和各类实体之间的复杂关系。本规范是在微软亚洲研究院《命名实体标注规范》4.0 版的基础上编制的。由于规范实际上涵盖了文本中词语和各类实体的标注规则与实例，所以更名为《中文文本标注规范》（Tokenization Guidelines of Chinese Text）5.0 版。

1.2 导读

规范的第一章（概述）、第二章（专有名词标注总则）、第六章（数字串标注总则）以及第九章（分词歧义消解细则）是每个标注人员必读的材料。其它章节收集了大量的实体标注规则与实例，用以补充各类实体定义的不足。凭借这些具有上下文信息的词例化实例可以进一步提高文本标注的精度和一致性，所以它们是供标注人员经常查阅的参考资料。诚恳欢迎读者对本规范和带标语料中的错误提出宝贵意见，以便及时更正。批评和意见请寄黄昌宁：cnhuang@msrchina.research.microsoft.com。

1.3 标注格式

format-1 是面向标注人员的格式：

/十月九日/上午/ → /*[dat 十月九日]*/*[tim 上午]*/

format-2 是基于 XML 的标注格式：

/十月九日/上午/ → <w><TIMEX TYPE= “DATE ” >十月九日</TIMEX></w><w><TIMEX TYPE= “TIME ” >上午</TIMEX></w>

TIMEX 是时间表达式，日期 *DATE* 和时间 *TIME* 是它的两个子类。

考虑到本规范主要是为标注人员编写的，以后的例子主要以第一种格式（format-1）表示。想了解更多 XML 格式的读者，请参见 MET-2 Guideline¹。

¹ MET (Multiple EntityTask) 是 1997 年第七届 MUC (Message Understanding Conference) 会议多实体识别任务的简称。MET-2 是当年美国 NIST 公布的命名实体标注规范。可查阅：http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/ne_task.html

1.4 命名实体标记集

表 1-1 是本规范定义的全部命名实体标记，包括专有名词（*NAMEX*）、时间表达式（*TIMEX*）、数字表达式（*NUMEX*）、度量表达式（*MEASUREX*）和地址表达式（*ADDREX*）等类五大类及其下属的三十个子类。

大类	子类	Format-1 标注集	Format-2 标注集
NAMEX	Person	P	PERSON
	Location	L	LOCATION
	Organization	O	ORGANIZATION
TIMEX	Date	dat	DATE
	Duration	dur	DURATION
	Time	tim	TIME
NUMEX	Percent	per	PERCENT
	Money	mon	MONEY
	Frequency	fre	FREQUENCY
	Integer	int	INTEGER
	Fraction	fra	FRACTION
	Decimal	dec	DECIMAL
	Ordinal	ord	ORDINAL
MEASUREX	Rate	rat	RATE
	Age	age	AGE
	Weight	wei	WEIGHT
	Length	len	LENGTH
	Temperature	tem	TEMPERATURE
	Angle	ang	ANGLE
	Area	are	AREA
	Capacity	cap	CAPACITY
	Speed	spe	SPEED
	Acceleration	acc	ACCELERATION
Other measures	mea	MEASURE	
ADDREX	Email	ema	EMAIL
	Phone	pho	PHONE
	Fax	fax	FAX
	Telex	tel	TELEX
	WWW	www	WWW
	Postalcode	pos	POSTALCODE

表 1-1 命名实体的标记集

1.5 基本原则

1.5.1 基本考虑

- 通用性：尽量遵循国际标准 MET-2 和 ER-99²，不同之处在本规范中阐明。
- 实用性：可用于 LSP (Lexical Service Platform), TTS (Text To Speech conversion), IR (Information Retrieval), IE (Information Extraction), QA (Question Answering), IME (Input Method Editor)等应用系统。

1.5.2 标注对象

1.5.2.1 词表词与未登录词

本规范认为：文本中的任何一个词要么是词表词 (LW)，要么是未登录词 (OOV)。两者都是文本的标注对象。未登录词可以进一步分成命名实体 (NE)、词法派生词 (MDW) 和新词 (NW) 等三部分。本规范定义的命名实体是未登录词的主体。

(1) 命名实体 (NE)

命名实体可以进一步分成如下五大类共三十个子类（详见表 1-1）：

专有名词 (NAMEX) 包括人名 (P)、地名 (L) 和机构名 (O) 等 3 种。

时间表达式 (TIMEX) 包括日期 (dat)、时间 (tim) 和时段 (dur) 等 3 种。

数字表达式 (NUMEX) 包括百分数 (per)、钱款 (mon)、频度 (fre)、整数 (int)、分数 (fra)、小数 (dec)、序数 (ord) 和比率 (rat) 等 8 种。

度量表达式 (MEASUREX) 包括年龄 (age)、温度 (tem)、角度 (ang)、长度 (len)、面积 (are)、容积 (cap)、重量 (wei)、速度 (spe)、加速度 (acc) 和其它 (mea) 等 10 种。

地址表达式 (ADDREX) 包括电子邮箱 (ema)、电话 (pho)、传真 (fax)、电报挂号 (tel)、邮政编码 (pos) 和网址 (www) 等 6 种。

在标注过的文本中，词的边界一律用斜线 (slash) 表示。除了词表词以外，每个独立的命名实体（即非嵌入到词表词内部的实体，见 1.5.2.3）也被视为一个词，其标注符号及形式详见本规范。

(2) 词法派生词 (MDW)

以词表词 AB 的重迭形式 AABB 和 AB/AB 为例：

/转轨/哪有/像/人/说得/那般/轻轻松松/?

/积累/多了/, /抽出/时间/, /认真/整理/整理/,

(3) 新词 (NW)

一个新词的左右两侧用符号 ‘&’ 标示，其内部的切分符保留³，如：

² ER 是实体识别 (Entity Recognition) 的简称。ER-99 是 1999 年 NIST 在 MET-2 的基础上修订的命名实体标注规范。

³ 为了便于处理，在公布给 SIGHAN Bakeoff-3 使用的语料中新词符号 ‘&’ 及其中的切分符号将被删除。这时文本中的新词将等同于词表词。

/&桑拿&/浴/
/天时地利&人/和&/;
/L 罗/货币/&列/伊&/

以下是一些真实的例句，例句中的实体标注符号请参阅表 1-1。

[Example-1]

/[dat 6 月 2 9 日]/、/[dat 3 0 日]/[tim 晚上]/，/[L 北京市]/下/了/[int 两场]/大雨/，/笔者/居住/的/宿舍/楼/前/，/宽/[len 六七米]/、/长/[len 3 0 多米]/的/路/上/积水/达/膝盖/之上/。

<sentence>

<w><TIMEX TYPE="DATE"> 6 月 2 9 日 </TIMEX></w><w>、 </w><w><TIMEX TYPE="DATE"> 3 0 日 </TIMEX></w><w><TIMEX TYPE="TIME"> 晚 上 </TIMEX></w><w>， </w><w><NAMEX TYPE="LOCATION">北京市</NAMEX></w><w>下 </w><w>了 </w><w><NUMEX TYPE="INTEGER"> 两 场 </NUMEX></w><w>大 雨 </w><w>， </w><w>笔 者 </w><w>居 住 </w><w>的 </w><w>宿 舍 </w><w>楼 </w><w>前 </w><w>， </w><w>宽 </w><w><MEASUREX TYPE="LENGTH"> 六 七 米 </MEASUREX></w><w>、 </w><w>长 </w><w><MEASUREX TYPE="LENGTH"> 3 0 多 米 </MEASUREX></w><w>的 </w><w>路 </w><w>上 </w><w>， </w><w>积 水 </w><w>达 </w><w>膝 盖 </w><w>之 上 </w><w>。 </w>

</sentence>

[Example-2]

/[dat 6 月中下旬]/，/笔者/到/[L 意大利]/、/[L 西班牙]/等/国/访问/时/，/一个/很/深/的/感受/是/[L 意]/、/[L 西]/两国/的/高速公路/非常/发达/，/东西南北/，/纵横/成/网/，/四通八达/。

<sentence>

<w><TIMEX TYPE="DATE"> 6 月中下旬 </TIMEX></w><w>， </w><w>笔 者 </w><w>到 </w><w><NAMEX TYPE="LOCATION">意大利</NAMEX></w><w>、 </w><w><NAMEX TYPE="LOCATION">西班牙</NAMEX></w><w>等 </w><w>国 </w><w>访 问 </w><w>时 </w><w>， </w><w>一 个 </w><w>很 </w><w>深 </w><w>的 </w><w>感 受 </w><w>是 </w><w><NAMEX TYPE="LOCATION">意 </NAMEX></w><w>、 </w><w><NAMEX TYPE="LOCATION">西 </NAMEX></w><w>两 国 </w><w>的 </w><w>高 速 公 路 </w><w>非 常 </w><w>发 达 </w><w>， </w><w>东 西 南 北 </w><w>， </w><w>纵 横 </w><w>成 </w><w>网 </w><w>， </w><w>四 通 八 达 </w><w>。 </w>

</sentence>

[Example-3]

/[O 县委]/决定/选派/任/了/[dur 八年]/[O 城建局]/长/的/[P 周欣光]/担任/[O 老干部局]/长/。

<sentence>

<w><NAMEX TYPE="ORGANIZATION">县委 </NAMEX></w><w>决 定 </w><w>选 派 </w><w>任 </w><w>了 </w><w><TIMEX TYPE="DURATION"> 八 年 </TIMEX></w><w><NAMEX TYPE="ORGANIZATION">城建局 </NAMEX></w><w>长 </w><w>的 </w><w><NAMEX TYPE="PERSON">周欣光 </NAMEX></w><w>担 任 </w><w><NAMEX TYPE="ORGANIZATION">老干部局 </NAMEX></w><w>长 </w><w>。 </w>

</w>

</sentence>

[Example-4]

/[L 喇嘛寺村]/地处/[L 承德避暑山庄]/,/[L 山庄]/寺庙/林立/,/僧侣/穿梭/,/[L 山庄]/[L 外八庙]/的/[ord 第一个]/庙/就/是/[L 喇嘛寺]/。

<sentence>

<w><NAMEX TYPE="LOCATION">喇嘛寺村</NAMEX></w><w>地处</w><w><NAMEX TYPE="LOCATION">承德避暑山庄</NAMEX></w><w> , </w><w><NAMEX TYPE="LOCATION">山庄</NAMEX></w><w>寺庙</w><w>林立</w><w> , </w><w>僧侣</w><w>穿梭</w><w> , </w><w><NAMEX TYPE="LOCATION">山庄</NAMEX></w><w><NAMEX TYPE="LOCATION">外八庙</NAMEX></w><w>的</w><w><NUMEX TYPE="ORDINAL">第一个</NUMEX></w><w>庙</w><w>就</w><w>是</w><w><NAMEX TYPE="LOCATION">喇嘛寺</NAMEX></w><w>。</w>

</sentence>

1.5.2.2 L, P, O, dat, tim, dur 等实体的边界允许跨越多个词表词

例如：

/[L 瑞典]/[O 斯德哥尔摩国际和平研究所]/
/[O 中国工商银行上海市分行]/
/[tim 下午当地时间5时59分]/

1.5.2.3 专名的标记（L, P, O）可以插入到一个词表词的内部

例如，词表词‘抗日战争’和‘事后诸葛亮’中的地名和人名应分别予以标注。

/抗[L 日]战争/	---正确标注。
/抗日战争/	---未标出‘L’，是错误标注。
/抗/[L 日]/战争/	---插入分词标记，是错误标注。
/事后[P 诸葛亮]/	

1.5.2.4 数字串（除专名以外的其他四类表达式）的标记不得插入到词表词的内部

1.5.2.4.1 dat, tim 等标记不得插入到一个词表词的内部

词表词‘夏令营’、‘春耕’、‘冬训’、‘早出晚归’中的‘夏、春、冬、早、晚’等词素都有‘dat’和‘tim’的意思，但不得标注。例如，

/[dat 冬]训/	---错误标注。
/[tim 早]出[dat 晚]归/	---错误标注。

然而词表词被整体标注为‘dat 和 tim’的情况是常有的，例如：

/[dat 初冬]/	----‘初冬’是词表词。
/[dat 夏季]/	----‘夏季’是词表词。
/告别/[dat 冬日]/的凝重、/[dat 春天]/的轻盈、/[dat 夏日]/的浪漫、	

- 注：在文本中具有比喻意义的‘春、夏、秋、冬’、历史上的‘今天、昨天、明天’不作标注。例如：

/[dat 今年]又/迎来/了/一个/科学/的/春天/
/“/在/陆地/资源/日渐/减少/的/今天,/ /
/他们的/明天/将/更加/辉煌/。/

1.5.2.4.2 int, ord 等标记不得插入到到一个词表词的内部

词表词‘五湖四海’、‘不管三七二十一’、‘三纲五常’中的数词不允许标注‘int’（整数）。例如，

[int 五]湖[int 四]海/ ----错误标注。
 /十年动乱/结束/不久/, / ----‘十年动乱’是词表词。‘十年’不标。
 /不管三七二十一/
 /三纲五常/

1.5.2.5 数词‘首、半、双、两’等

1.5.2.5.1 序数词素‘首’

词表中有许多词含有词素‘首’，如‘首创、首倡、首选、首发、首航、首飞、首演、首映、首战、首展、首席代表、首席科学家、首席执行官、首富、榜首、魁首、居首’等。但不可把词表词中的词素‘首’单独作为‘ord’（序数）来标注。

/首席执行官/ ----正确标注。
 /[ord 首席]执行官/ ----在词表词中插标‘ord’是错误的。

以下的词表词属于“首+量词”结构，可以整体作为‘ord’标注。例如：

[ord 首届], [ord 首次], [ord 首批], [ord 首位], [ord 首例]

1.5.2.5.2 分数词素‘半’

词表中有许多词含有词素‘半’如‘半价、半票、半饱、半身、半世、半辈子、上半时、下半场、半边’等，但不可把上述词表词中的词素‘半’标注为‘fra’（分数）。

/下半场/比赛[0 中国队]/未进/一球/
 /上半时/
 /下[fra 半]场/ ----在词表词中插标‘fra’是错误的。

以下的词表词可作为不同的数字串（dur, tim, fra, int, age）标注：

[dur 半年], [dur 半天], [tim/dur 半夜], [fra 半个], [int/age 半百]

- 注：半个西瓜中的‘半个’，与‘四半’中的‘半’概念不一样，前一个‘半’是指二分之一，后一个‘半’是量词，所以标注也不同！！

/[int 一个]/西瓜/分为/[int 四半]/
 /[fra 半个]/西瓜/

1.5.2.5.3 整数词素‘双’

当数词‘双’成为词表词的一个词素时，如“双方、双边、双手、双打、双杠、双轨、双层、双目、双亲”等，一律不作为整数(‘int’)标注。对于非词表词，只标[int 双]。例如：

/窗外/又起风/了/, /双层的/窗/硬是/阻挡/不住/沙尘/的/侵扰/。/
 /双方/认为/, /[L 中][L 美]/两国/应该/从/战略/的/高度/和/长远/的/角度/

- 注：一般情况下，数词和“方”之间不切分整体标为‘int’。但“四方”是词表词所以不标。

[int 三方]/会谈/
 /举行/四方/会谈/

以下是相关的例子：

/用/任何/一/部[int 双]/音频/电话/只需/拨打/[pho 2 5 8 0]/就可以
 /部队/进行/的/海上/训练/、/[int 双]/机/穿云/、/超低空/等/高难/课目/训练/
 /全村/[are 7 0 0 亩]/旱地/都种上/了/[int 双]/膜/棉/, /

1.5.2.5.4 整数词素‘两’

当数词“两”成为词表词的一个词素时，如“两国、两会、两地、两者、两头、两手、两边、两旁、两侧”等，一律不作为数位串（‘int’）标注。例如：

/使/两国/的/友好/合作/得到/巩固/和/发展/。

- 注：一般情况下，数词和“国”之间是要切分的，如：*[int 五]/国/元首/前/些年/我/对/参加/【/两会/】/总是/有点/发怵/。/*
/大街/两旁/店铺/林立/
/戏台/两侧/立柱/上/有/这样/一/副/对联：/
/中间/[int 两间]/是/客厅/，/两边/是/卧室/和/书房/。/
/对/分散/居住/的/“/五保/”/户/，/镇/、/村/[int 两]/级/拨/专款/
[int 两]/车/饮料/以及/办公/桌椅/，/
- 注：临时量词“车、船、床、桌、屋子、院子”等不进入‘int’标注。
/成为/[O 议会]/[int 两]/院/审议/的/重点/和/舆论/关注/的/焦点/
- 注：‘两院’不是词表词，所以应当切分标注如上。
/及早/进行/政治/谈判/推动/[L 两岸]/关系/发展/
/沿江/[int 两]/岸/苗家/吊脚楼/上/的/观众/
- 注：词表词‘两岸’是专指台湾海峡两岸的地名。如果泛指江河两岸，则不作为地名标注，而且要切分并标数词“两”为‘int’。
/一下/进/了/[int 两]/球/

1.5.3 基本规定

1) 标注时，不得在原来的文本中加入回车换行符。

2) 对于 NIST 制定的两个中文 NE 标准：MET-2 和 ER-99。前者已有系统参加评测，它们的评测结果可供后来者参考；后者是前者的修订版，但尚未有系统参加测试。本规范与这两种标准不同之处将尽可能在注释中加以说明。例如：

/[dat 去年上半年]/

- 注：MET-2 把‘去年上半年’整体视为‘dat’；ER-99 则只将‘上半年’视为‘dat’。

3) 对于微软研究院根据自己的需要而加入的标记，本规范也将在注释中加以说明。比如本规范要求的如下标注：

/[P 邓小平]/理论/

- 注：MET-2 和 ER-99 规定，‘理论，主义，思想，定律’等词前面的人名均不作为专名标注（见 2.8）。

第二章 专有名词标注总则

2.1 专有名词（NAMEX）标注通则

对于人名、地名和机构名这三类专有名词，MET-2 和 ER-99 之间的差异甚微，在它们给出的示例中，只有两处不同：‘*中南美*’和‘*长江流域*’（具体情况见后）。所以在制订人名、地名、机构名的标注规范时，我们没有刻意去区分这两个标准，而是力图把它们统一地融入本规范。

下面给出人名、地名、机构名的定义。

2.2 专有名词是具体的、特定的，而不是抽象的、泛指

比如：‘*上苍*、*老外*、*姑娘*、*小镇*、*企业*’等就不应视为专有名词。

2.3 复合专有名词的标注不允许嵌套

在 MET-2 和 ER-99 标准中，任何命名实体都不允许嵌套。换句话说讲，只标一个实体的最长边界，不标其内部包含的其它实体。

2.4 人名、地名、机构名中的数字串不单独标出

例如：

/[P 龟山一郎]/

/[L 德富路二四一至二六三号]/

/[O 北京 101 中学]/

/[O 北京[ord 四]中]/

----这种嵌套式的的标注是错误的！

2.5 含有外文和数字的命名实体应整体一起标注

例如：

/[O American 航空公司]/

/[O SONY 公司]/

2.6 当两个实体用虚词‘的’连接时应分别标注为两个实体

例如：

/[L 美国]/的/[L 纽约]/

/[L 美国]/的/[P 理查德本森]/

但当‘的’成为实体的一部分时，要整体一起标注。例如：

/[O 美的电器集团]/

2.7 实体前后有引号或书名号的情况

如果一个命名实体中间有引号或书名号，则引号或书名号是该实体的一部分。如果一个实体被外面的引号或书名号括起来，那么其引号或书名号就不作为实体的一部分标注。例如：

/[O “阿克布拉克”中哈合资企业]/
/[O 美国《幸福》杂志]/
/[O 星岛日报]/的/社论/说/

2.8 短语内部包含实体、但整体又不是命名实体的情况

ER-99 规定：如果一个短语内部包含实体、但整体又不是命名实体，则一律不作标注。本规范则要求对该短语中的实体部分加以标注。例如：

/[L 巴拿马运河]/条约/

- 注：ER-99 认为，‘巴拿马运河条约’整体不能分解，其中的地名不应标注。但本规范把其中的‘巴拿马运河’单独标为地名。

/[L 巴拿马运河]/ ---- ‘巴拿马运河’单独出现时，作为地名标注。
/[L 香港]/脚/

- 注：英文为“Hong Kong foot”，类似于“athlete’s foot”，不可分解，所以 ER-99 规定整体不标。本规范，仍将‘香港’标为地名。

/[L 美国]/小姐/

- 注：原文为“Miss America”，指选美活动中获全美第一名的小姐。对此 ER-99 规定整体不标。本规范，仍将‘美国’标为地名。

/[L 美国]/姑娘/ ---- ER-99 对本例的‘美国’也是标注的。

/[ord 第四十六届]/[O 太平洋亚洲旅行协会]/年会/

- 注：此例在 ER-99 中整体不标，理由是“不可分解”。本规范认为找不出充分理由说明其不可分解。所以我们将‘太平洋亚洲旅行协会’标为机构名。‘第四十六届太平洋亚洲旅行协会年会’整体不是机构名。

/[P 毛泽东]/思想/

/[P 马克思]/主义/

/[P 马克思]/主义/ ---- 错误标注！因为‘马克思主义’是词表词。

/[P 阿佛加罗]/定律/

- 注：ER-99 规定，在‘理论、主义、思想、定律’等词前面出现人名时，是整体不可分解的字符串；因此该字符串和其中的人名都不标注。但本规范仍将标注其中的人名。

2.9 与军队相关的情况

当泛指某个国家的军队（如‘英军、美军’等）时，不是机构名；当指一个具体的军种（如‘空军、陆军、海军’等）时，要标注为机构名。例如：

/[L 美]/军/飞机/

/[O 斯里兰卡空军]/

/[O 英国皇家空军]/

但是，有如下特殊情况：

/[L 济南军区]/ ---- 军区是 ‘L’ 而不是 ‘O’。

/[L 彼得森空军基地]/ -----军事基地是‘L’而不是‘O’。
/[L 西非]/&维/和&/部队/ -----‘部队’不作为机构名标准。

2.10 多媒体、产品和条约中的人名、地名、机构名

ER-99 规定：当人名、地名、机构名属于多媒体、产品和条约时，均不加标注。但本规范对上述实体名还是要标注的。例如：

/[P 邓小平]/-/[片]/的/播出/

- 注：ER-99 规定，电视节目的名字‘邓小平’不标。本规范仍把它标为人名。此外，‘邓小平’作为片名，在规范的文本中应当用书名号括出，如《邓小平》。

/二战/ -----‘二战’是事件，所以不标注。

/[L 香港]/百/题/今天/为/您/解答/

- 注：ER-99 规定，‘香港百题’是电视片的标题，所以专名‘香港’不予标注。但本规范仍把‘香港’标为地名。下面其它的例子就不一一解释了。

/这/本/介绍/[P 毛泽东]/的/小说/ -----‘毛泽东’要标注。

/这/本/名/为/[P 毛泽东]/的/小说/ ----- ER-99‘毛泽东’不标。

/[L 广州]/条约/ ----- ER-99‘广州’不标。

/[L 辽][L 沈]/战役/ ----- ER-99‘辽沈’不标。

本规范在后面还要对人名、地名、机构名中不加标注的情况作专门的说明，详见下面的各章节标注细则。

2.11 别名或简称的标注

对人名、地名、机构名的别名或简称要标注。例如：

/[O IBM]/

/[L 深]/[L 沪]/股市/

/[O 北约]/

/[L 中][L 美]/首脑/互访/

/[L 中]/[L 文]/双方/一致/认为/

- 注：由于‘中美’是词表词，标注地名时不可插入分词标记。‘中文’也是词表词，但这里是指‘中国’和‘文莱’，所以标成地名时需要在两个简称中插入分词符号。这样的词表词还有‘中意、意中、中巴、日中、中肯、中非’等。‘巴中’是一个地名，但表示‘巴基斯坦’和‘中国’时需要用分词符号把两个简称分隔开。

- 注：对于简称中嵌套的人名、地名、机构名不予标注，如：

/[O 中共]/ -----‘中’指‘中国’，但不标。

/[O 中共中央政治局]/ -----同理，不标注‘中’。

第三章 人名

人名一般包含‘姓’和‘名’两部分，‘姓’是表明家族的字，有单姓和复姓之别；‘名’也就是名字，是一种称号，由一个或几个字组成，跟‘姓’合在一起，用来代表一个人，以区别于别的人。下面将对人名的标注规则进行详解。

3.1 人名标注规则

正常情况下，人名一般包含‘姓’和‘名’两部分，标注规则如下表所示：

序号	情况	标记方法	例子
1	只含 姓，没有 名	标出姓氏部分	[P 庄]、[P 欧阳]、[P 司马]
2	只包含名字	标出名字部分	[P 育焜]
3	姓名	姓名整体标出	[P 苏宗哲]、[P 萝莉胡吉温]
4	姓名 姓 名 + 称谓 称谓 + 姓名 姓 名	只标出： 姓名 姓 名	[P 许]/乡长 [P 孙宝钿]]/局长 [P 志航]/教授 县长/[P 阮刚猛] 称谓包括：(副) 主席、(副) 总理、(副) 部长、(副) 厅长、(副) 局长、(副) 股长、(副) 校长、(副) 教授、老师、(副) 班长、博士、硕士、(副) 研究员、高工、董事长、董事、(副) 市长、课长等
5	前缀 + 姓名 姓 名 姓名 姓 名 + 后缀	只标：姓名 姓 名	大[P 李] 小[P 李] 老[P 李] 大[P 李明] [P 李]/总
6	姓名 + 姓名	分开来标	[P 李向东]/[P 李向阳]
7	外国人名	作为一个整体来标	[P 罗马里奥] [P 马拉多纳] [P 比尔·盖茨]

*****说明：当人名中包含‘·’时，整体标注为人名，如[P 比尔·盖茨]。

3.2 人名标注细则

3.2.1 人名的示例和详细说明

3.2.1.1 人名实例

/[P 颜惠忠]/

/[P 连战]/

/[P 凡·高]/

/[P 陈方安生]/

---当妻子与丈夫的名字写在一起时，要作为一个人名整体标注为 P!

3.2.1.2 称谓、绰号、官职不作为人名的一部分

称谓、绰号、官职(如‘先生、总理’等)不作为人名的一部分。例如，

/[P 张]/经理/

/[P 李]/市长/

/[P 陈]/姓/游客/说/

/[P 刘]/[ord 二]/嫂/

/[P 周]/总理/

/[P 雷锋]/同志/

/[P 奥尔布赖特]/国务卿/

3.2.1.3 当称谓和姓名不可分时应整体标注为人名

/[P 李主席登辉]/先生/

/处/[P 李犯清龙]/死刑, /

/[P 李犯×龙]/持/刀/行凶/杀害/无辜/青年/, /

3.2.1.4 ‘几世’、‘几代’ 要作为人名的一部分

/[P 十四世达赖丹增加措]/

/[L 英国]/女王/[P 伊丽莎白二世]/

3.2.1.5 家族实体

/[P 蒋]/氏/父子/

/[P 西迪]/兄弟/

3.2.1.6 圣人和宗教人物要标注为人名

/[P 释迦牟尼]/

/[P 达赖]/喇嘛/

3.3 虚构的人物、动物的名字要标注为人名

3.3.1 在童话、小说中虚构人物要标注为人名

/[P 孙悟空]/
/[P 玉皇大帝]/

3.3.2 虚构的动物和非人的人物要标注为人名

/[P 唐老鸭]/
/[P 花仙子]/
"/[P 盼盼]"是国内外著名的熊猫明星，
争相目睹狮王[P 木法沙]和王后[P 色拉碧]产下的小王子[P 辛巴]
走进一家饭馆，发现老板就是大灰狼[P 罗克]。/

3.3.3 用称谓或朝代等名号来指称特定人时要标注为人名

例如：

/[P 康熙]/
/[P 乾隆]/
/[P 秦始皇]/
/[P 老子]/
/[P 孔子]/

3.4 不标注为人名的各种情况

3.4.1 虚构的非人的植物的名字不作为人名标注

如：

"/彩霞，石子/小声/嘟哝/着，"/多恶心的名字！"/
/电磨/姐姐/故意/气气/小毛驴，/说："/输/了/，/可/不能/哭鼻子/。"/
"/卡车/哥哥，/我/和/你/来/比/一/比/谁/运/得/多/，/怎么样？"/
/好像在说："/荷花/姐姐，/你好！/

3.4.2 对于嵌套在地名和机构名中的人名，不作标注

如：

/[L 嘉诚广场]/
/[O 中山大学]/
/[O 宋庆龄基金会]/

3.4.3 作为书名或画名的人名

作为书名或画名的人名 ER-99 不标（见 2.8），但本规范是要作标注的。如：

/世界/名画/《/[P 蒙娜莉萨]/》/
/《/[P 蒋介石]/与/[P 毛泽东]/》/

3.4.4 法律、法庭事件、天气形成、疾病和奖金等五种情况

当人名后面紧跟法律名、法庭事件、天气形成、疾病、奖金这五种情况时，人名不标注。

例如：

/里/氏/[ord 六点二级]/	-----‘里’不标。
/专家/呼吁/人们/要注意/沙/氏/杆菌/	-----‘沙’不标。
/[P 诺贝尔]奖/	----- ER-99‘诺贝尔’不标。

3.4.5 在人名后面出现‘基金会’时要整体标注为机构名

/[O 李嘉诚基金会]/

所以‘基金会’和‘奖、奖金’是不同的两种情况，需加以区别。又如

/[O 李嘉诚股份有限公司]/
/[O 诺贝尔股份有限公司]/

第四章 地名

地名包括洲、海洋、国家、省、市、县、地区、街道、乡、镇、村、机场、军事基地、军区、铁路、公路、桥梁、海峡、海湾、港湾、河流、湖、公园、草原、煤矿、牧场、养殖场、音乐厅、剧院、教堂、寺庙、图书馆、博物馆、美术馆、展览中心、公园、动物园、植物园、火车站、广场、大厦、大楼、体育场(馆)、游泳馆(池)、赛车场、商城、超市、书店(城)等城市公共设施，还包括某些特定的城市建筑和虚构的处所。详见下表。

4.1 地名标注规则

序号	情况	标记方法	例子
1	只是单独地名	标出地名部分	[L 中国] [L 竹塘乡]
2	地名 + 地理(行政)单位	作为整体标出	[L 北京市] [L 台北县] 地理单位如：省、地区、市、县、乡、镇、村、店、庙、沟、屯、坟、崖、海洋、河、川、江、峡谷、海湾、港湾、丘陵、湖、半岛、三角洲、区、街、路、街、街道、社区、小区、公寓、音乐厅、剧院、图书馆、博物馆、美术馆、展览馆、公园、动(植)物园、火车站、广场、大厦、大楼、体育场(馆)、游泳馆(池)、赛车场、商城、超市、书店(城)等城市公共设施及象征性建筑物、军事基地、军区等。 [L 天安门广场] [L 艾菲尔铁塔]
3	包含上、下位的地名（即合成地名）以及并列的地名	一律分别单独标出	[L 山东省]/[L 青岛市]/[L 胜利广场] [L 青岛市]/[L 孙中山广场] [L 北京市]/[L 海淀区]/[L 知春路]/[L 希格玛大厦] [L 北京]/、/[L 天津]/、/[L 上海]
4	地名简称	单独标出	[L 鲁]/、/[L 冀]/、/[L 京]
5	并列的简称	单独标出	[L 中]/[L 俄]/两国/领导人/进行/了/会晤 [L 港]/[L 澳]/[L 台]/地区
6	地名包含人名以及地名包含地名的情况	地名中的人名、地名不标	[L 李嘉诚广场] [L 南京路]
7	地名 +地名关键词表达一个完整的概念时	相对完整的地名	[L 南非共和国] [L 宁夏回族自治区] [L 香港特别行政区]

4.2 地名标注细则

4.2.1 地名实体示例

/[L 北京]/

/[L 亚洲]/

/[dat 2008 年]/[L 奥]运会/

/[L 中国]人/

/[L 中国]人民/

----‘中国人、中国人民’都是词表词。

/[L 朝鲜]/南北/对话/

----不标注‘南’、‘北’。

- 注：词表词“京剧、京白、京腔、京味儿”中的“京”字要标注为：

/[L 京]剧、/[L 京]白、/[L 京]腔、/[L 京]味儿/

/[L 台东火车站]/

/[L 阜南文化公园]/

/[L 基隆文化中心广场]/

/[L 高雄港第一港口]/

/[L 苏澳镇]/[L 南方澳渔港]/

/环/[L 渤海湾]/地区/的/天然气/市场/

/来自/[L 沈阳军区]/各/集团军/

/[L 梅狮路]后段/

/[L 中横公路]天祥段/

/[L 华禄溪]/及/[L 碧绿隧道]/

/[L 南二高]/[L 高雄支线]/

/[L 台廿一线]/

/[L 美国空军基地]/

/[L 上海]/[L 国际航运大厦]/

/[L 上海]/[L 虹口足球场]/

/[L 上海博物馆]/

/[L 上海]/[L 城市规划展示馆]/

/[L 石家庄]/[L 富强电力新村]/

/[L 西安第二长途通讯大楼]/

/[L 北京市]/[L 王府井百货大楼]/

/[L 广深铁路]/以及/[O 深圳发展银行]/部分/高官/也/被/免职/

/[L 汉江]/上/的/[L 圣水大桥]/

/[L 新亚欧大陆桥]/

---从世界知识知道此处大陆桥的名字叫‘新亚欧大陆桥’，是不可分解的。

4.2.2 地名指示词（如‘国、省、市’等）视为地名的一部分一起标注

地名指示词（如‘国、省、市’等）视为地名的一部分一起标注。复杂的、具有包含关系的地名要分开标注，但分开标注时不可把一个有完整意义的地名拆散。以下是正确的标注：

/[L 德国联邦]/政府/总理/
 /[L 基隆市]/
 /[L 台东县]/
 /[L 南山部落]/
 /[L 美国]/[L 马里兰州]/
 /[L 约旦河]/
 /[L 朝鲜半岛]/
 /[L 长江三角洲] -----‘长江三角洲’是词表词。
 /[L 吉林省]/[L 延边朝鲜族自治州]/[L 图们市]/

以下两例均为错误的标注，因为‘延边朝鲜族自治州’是具有完整意义的地名：

/[L 吉林省]/[L 延边]/[L 朝鲜族自治州]/[L 图们市]/
 /[L 吉林省延边朝鲜族自治州]/[L 图们市]/

- 注：在 ER-99 的标准测试集中，把‘中国西昌卫星发射基地’整体标为地名。我们认为这是错误的，因为在一个地名中不应当包含具有上、下位关系的另一地名。正确的标注是：

/[L 中国]/[L 西昌卫星发射基地]/
 /[L 美国洛克希德·马丁卫星测控中心]/和/[L 中卫公司测控站]/
 /从/[L 法]/属/[L 圭亚那]/[L 库鲁航天中心]/发射/

- 注：本规范不采用 ER-99 的标注：‘[L 法属圭亚那库鲁航天中心]’。

/[L 武汉]/[L 长江大桥]/
 /[L 上海]/[L 中山公园]/

- 注：尽管其它城市也有‘长江大桥’和‘中山公园’，但在当地它们已构成完整的地名，所以应单独标注。

/位于/[L 朝阳门]/外/商务/区/之中，/
 /[L 盛华公寓]/坐落/于/[L 西直门]/内/[L 冠英园小区]/

- 注：‘内、外’都不在标注范围之内，但如果地名中的‘内、外’去掉不能说明是一个完整的地名时，‘内、外’要标注在地名内。如：

/[O 外交部]/位于/[L 北京市]/[L 朝阳门内南小街52号]/
 /[L 西直门外大街71号]/

4.2.3 并列的地名应分别标注

对于并列的多个地名应分别标注。对于嵌套在地名中的人名、地名和机构名不再单独标注。例如：

/[L 中]/[L 意]/双方/ -----‘中意’是词表词，作为国名时要切开。
 /[L 香港]/和/[L 澳门特别行政区]/
 /目前/已/[int 1 2个]/[L 中]/、/[L 东欧]/国家/
 /[L 北京]/[L 上海]/
 /[L 科]/[L 伊]/边境/

4.2.4 跨国家的和国家内部的地名

/[L 西非]/国家领导人/
 /从/[L 陕]/[L 甘]/革命/老区/到/沿海/经济特区/, /
 /[L 亚太]/ ---- ‘亚太’是词表词，它是一个地名，而不是两个地名。
 /[L 近东]/和/[L 北非]/

4.2.4.1 表示地理方位的名词

一些表示地理方位的名词如‘南半球、北半球、江南、江北、西南、西北、华南、华北、华中、东北’等虽然不完全具备确指性，也要作为地名标注为 L。

/[L 汉水]/流域、/[L 西南]/地区/东部/
 /[L 江南]/大部、/[L 华南]/有/小/到/中雨/
 /近/[dur 两天]/造成/[L 东北]/、/[L 华北]/地区/的/降雨/天气/系统/
 /迫使/[L 北半球]/的/副热带/高压带/在/[L 青][L 藏]/地区/“/断裂”/

- 注：上述地名后面的方位词‘南部、北部、东部、西部’不应包括在地名的括号里，因为其所指的区域是更不确定的。

4.2.4.2 方位词修饰地名实体时要整体标注为‘L’

/[L 东西九龙]/ ----这是一个并列的地名。
 /一代/又/一代/海/测/官兵/犁/波/耕/浪/于/[L 南中国海]/,
 /[L 北爱尔兰]/
 /[L 中西伯利亚]/

- 注：ER-99 将此例标为‘中/[L 西伯利亚]’。我们认为它整体是一个专指性的地名。

/[L 中南美]/

/[L 东南亚]/

- 注：ER-99 要求把上面两个地名分别标注为‘[L 中]/[L 南美]’和‘[L 东]/[L 南亚]’。其实‘中南美’指‘中美’和‘南美’两个地名，而‘东南亚’是一个地名。这样的细节需要专门的地理知识才能做出判断。所以我们不遵循 ER-99 的这条规则。

4.2.5 地名实体受时间词修饰时，时间词不标

/前/[L 苏联]/
 /前/[L 南]/地区 ---- ‘南’指南斯拉夫，时间词‘前’不标。

4.2.6 只有经纬度在一起时才能标注为‘L’

只有经纬度在一起时才能标注为‘L’，否则经度或纬度单独标为角度‘ang’。如：

/震/中/位于/[L 北纬三十六点二零度，东经九十点二九度]/
 /并/将/卫星/定点/在/[L 东经 110.5 度赤道]/上空。/
 /震/中/位于/[ang 北纬 30.5 度]/，/

4.2.7 天体的标注

/[L 宇宙]/
 /[L 地球]/
 /[L 太阳]/
 /[L 太阳系]/
 /[L 银河]/
 /[L 银河系]/
 /[L 月亮]/
 /[L 海王星]/
 /[L 东方红三号]/
 /[L “鑫诺1号”卫星]/

- 注：火箭只是卫星的发射工具，故火箭型号不作为星体标注。

/[dat 96年2月15日]/长征[ord 三号乙]/火箭/发射/失利，
 /长[ord 二]/捆/火箭 ----全名为“长征二号捆绑式运载火箭”。

4.3 不作地名标注的示例

/[L 阿]/[L 以]/冲突/

- 注：ER-99 和 MT-2 认为‘阿（阿拉伯）’不是一个特定国家的简称，本规范不采纳他们的规定。

/回答/[L 中]外/记者/的/提问/ ----‘外’不标。

4.3.1 ‘地区’一般不作为地名的一部分标注

仅当‘地区’特指行政单位时，才被视为地名的一部分。一般情况下，‘地区’泛指一片地方，不是地名的一部分。若不能确定时，‘地区’不作为地名的一部分标注。

/[L 港][L 澳][L 台]/地区/ ----‘港澳台’是词表词。
 /[L 巴尔干]地区/
 /[L 临沂]地区/现/更名/为/[L 临沂市]/

4.3.2 ‘平原、山脉、山区、盆地、沙漠、流域’不在标注范围内

‘平原、山脉、山区、盆地、沙漠、戈壁、流域、故里、故居、纪念馆、风景区、开发区、经济区’等都不在地名标注范围内。但当‘某某故居、故里、纪念馆’成为一个对外开放的旅游景点时，才作为地名标注。如：

/[L 云][L 贵]高原/
 ----‘云贵高原’是词表词不可分割，但‘云、贵’要分别标注 L。
 /[L 成都]平原/
 /[L 秦岭]山/脉/
 /[L 秦][L 巴]山区/

/[L 四川]/盆地
 /[L 撒哈拉]沙漠 ----‘撒哈拉沙漠’是词表词。
 /[L 长江]/流域
 /[L 毕加索故居]
 /造型/典雅/毗邻/[L 青云岩]/风景区/及/[L 北山湾]/旅游区/
 [L 约旦河西岸] ----因为‘约旦河西岸’是专指。
 /[L 海峡两岸]/ ----指‘台湾海峡两岸’。
 /[L 两岸]/

- 注：词表词‘两岸’只有在表示‘台湾海峡两岸’时，才作为地名标注为 L，当作为‘江河、湖泊’的两岸时，‘两岸’要切分标注。如：

/[L 长江]/的/丰姿/和/[int 两]/岸/的/美景/尽收眼底。/
 /祖国/[L 大陆]/

- 注：‘内地’虽然指中国大陆，但不作为地名标注，这里遵从了 ER-99 的规定。‘特区’只有在确指是‘香港’和‘澳门’时才作标注。如：

/来自/内地/和/[L 香港特区]/
 /[L 特区]/政府/和/[L 香港]/同胞/正/以/喜悦/的/心情/
 /[L 中国]/[L 厦门]/经济特区/

4.3.3 对语言文字前的单音节地名不标，双音节的地名标注为 L

英语 ----对‘英’不标注。
 汉语 ----对‘汉’不标注。
 中文 ----对‘中’不标注。
 /对/[L 西藏]/地区/的/藏语/广播/
 /主张/台语/在/[L 台]/
 /用/[L 四川]/话/ ----如果‘语、文’前面的地名为双音节时，就要标注。
 /[L 荷兰]/语/

4.3.4 以‘族’或‘裔’结尾的词组中地名也要标注

MT-2 和 ER-99 规定：以‘族’或‘裔’结尾的词组中的地名不标注。因此‘华裔’、‘汉族’中的‘华’和‘汉（指汉族）’都不作为地名标，但‘华人、华侨、华裔、中医、中草药、中餐馆、亚运会、奥运会’里的‘华、中、亚、奥’仍需标注‘L’。本规范不采用这一规则。作为民族的名字，单音节的不标，双音节的标‘L’。

下面是一些标准实例：

/[L 美]籍/[L 华]人 ----“美籍华人”是词表词。
 /目的是/促进/[L 塞浦路斯]/西/族/与/土/族/的/和解
 /她/和/同是/[L 日]/裔/[int 三]/世/的/男/友/
 /通过/在/[L 中]医药/宝库/里/寻找/线索/
 /人们/纷纷/拥向/[L 中]餐/馆/，/一/时间/人/满/为/患/
 /[L 吉普赛]/人/ ----‘吉普赛’不是词表词。
 /[L 印地安]/民族/；/ ----‘印地安人’是词表词。

第五章 机构名

机构名包括：股票（证券）交易所、国家或国际组织、商业团体（公司、企业、工厂）、电视台、广播电台、报刊杂志、出版社、政党或党派、学校、科研院所、医院、诊所、邮电局、乐队、体育运动队、联盟、议会或代表大会、军队、咖啡厅、酒吧、饭店、旅馆，以及虚构的机构等。

5.1 机构名标注规则

机构名的后缀应视为机构名的一部分。

序号	情况	标记方法	例子
1	普通名字 + 机构名	整体标出	[O 板桥市胜捷公司]
2	地名 + 机构名	机构名整体标出	[O 北京市电信局] [O 台北县立莺歌高职] [O 台北看守所] [O 基隆长庚医院] [O 东直门敬老院] 机构名的关键词如： 幼儿园、各级学校、科学院、部委、实验室、 工厂、公司、报刊杂志、出版社、大使馆、领 事馆、咖啡店、快餐店、饭店、酒店、旅馆等
3	人名 + 机构名	机构名整体标出	[O 李嘉诚基金会]
4	简称	一律整体标出	[O 北约] [O 上轮集团] ----指上海轮胎集团 [O 白宫]/官员/表示

5.2 机构名标注细则

5.2.1 机构名标注实体示例

/[O 国防部]/长/[P 迟浩田]/
/[O 美国国防部]/长/[P 佩里]/
/[O 台北县地政局地权课]/
/[O 地政局]/
/[O 政风室]/接/获/检举/调查/
/[O 国军北投医院]/
/[O 三重地政事务所]/

/O 台湾银行宜兰分行/
 /O 省立关山工商/
 /O 基隆市光隆家商/
 /O 东信国小/
 /O 安乐国中/
 /O 原住民委员会/
 /O 连萧全国竞选总部/
 /O 北京钓鱼台国宾馆/
 /L 浙江/O 温州大酒店/
 /O 松下电工株式会社/
 /O 公司/英文/名称/O HUNAN FORE SCAPE TECHNOLOGY CO., LTD/
 /O 朝鲜人民武装力量部/副/部长/
 /O 美国海军/
 /O 欧共体/
 /O 中国国家生育委员会/
 /O 中国奥林匹克队/
 /O 披头四/
 /O 飞虎队/
 /敢死队/ -----泛指不标。

/但是/O 共和党/人/说/
 /O 土耳其议会外交关系委员会/
 /O 终战 50 周年国会议员联盟/
 /记者/来到/O 中山医科大学第一附属医院住院部/
 /O 中共中央政治局/常委、/O 中央纪委/书记/P 尉健行/

- 注：‘中国共产党’的简称‘中共’或‘共’要标注为‘O’。例如：

/ord 第二次/O 国/O 共/合作/

- 注：类似的简称‘党’，由于专指性不强，不标，如：

/但/这种/现象/的/产生/，/是/同/党/和/国家/尊师重教/的/方针/背道而驰/的/，

/全国/“/[dat 三八]/“/红旗手/、/全国/优秀/共青团员/

- 注：“三八红旗手”是词表词。但如果“三八”在文中被双引号断开，就要单独表为‘dat’。另外，词表词‘共青团员、共产党员、共产党人’，中的机构名不确指，所以一律不标。

/O 中共中央政治局常委会/

- 注：‘常委会’可以是机构名，‘常委’则不是。

/党/的/O 十四大/以来/

- 注：中共的‘X中全会’不是机构名，除了词表词‘三中全会’什么也不标以外，数词‘X’应单独标注为‘ord’。例如：

/根据/党/的/[ord 十五届]/[ord 二]/中/全会/

/O 八届全国人大/代表/P 陈妙珍/

/O 西藏政协/委员/强调/，/必须/旗帜/鲜明/地/反对/民族/分裂/

[O 澳门中华总商会]/会/董/兼/O 青年委员会/副/主任/

/O 足协/杯赛/冠军/O 北京国安队/ -----‘杯赛’是词表词。

/O 以国家电视一台/ -----指以色列国家电视一台

/[L 汉城]/[O 路透]/电/
 /前/[L 苏联]/[O 切尔诺贝利核电站]/泄漏/事件/
 /参加这次/比赛/的/还有/[O 日本]/、/[O 俄罗斯]/、/[O 美国]/、/[O 德国]/
 和/[O 意大利队]/。/
 /前往/[O 解放军驻港部队总部]/慰问/驻军/
 /[O 第四届和平小天使台湾访问团]/抵达/[L 重庆直辖市]/
 /[O 塔里班]/部队/已经/到达/[P 杜斯塔姆]/将军/的/家乡/
 /用/公款/购买/[O 靖国神社]/和/[O 护国神社]/的/祭祀品/
 /纪念币/正面/是/由/[O 解放军]/军徽/光，/八一南昌起义/和/[O
 解放军]/[O 陆]/[O 海]/[O 空]/三军/战士/的/图案/

- 注：词表词‘八一南昌起义’是一个事件，不是机构名。‘三军’是词表词，所以数字‘三’不作为‘int’标注。
- 注：股市报导中的企业和公司名不论其前后有没有外文字符，一律作为一个整体标注成‘O’。例如：

/[O ST辽物资]/[dec 14 · 14 1]/[O 宁波中百]/[dec 20 · 35 4]/
 /[O DR沪港机]/[dec 11 · 19 4]/[O 鲁北化工]/[dec 8 · 05 1]/

- 注：商城或百货公司本应标注为‘L’，但作为股市中时企业时应标注为‘O’。
- 注：股票指数在没有明确说明是多少元的情况下一律标注为‘int’或‘dec’。
- 注：被命名的轮船、飞机、机车应标注为‘O’。例如：

/却/购/回/了/[int 3张]/[O “长月”号轮船]/船票/，/
 /[O 泰坦尼克号游轮]/上/的/这/对/情人/实在/浅/得/很。/
 /[O 美国“哥伦比亚”号航天飞机]/上/的/宇航员/

5.2.2 机构名的后缀是机构名的一部分

机构名的后缀是机构名的一部分，即要准确的标出机构名的最长边界（机构名的全称）。机构名中可以包含人名、地名和机构名，但对于它们不再单独标注。例如：

/[O 苗栗县环保局]/
 /[O 卫生署桃园医院]/
 /[O 兰阳民生医院]/前身/为/[O 吴外妇科]/
 /[O 台北爱乐青年管弦乐团]/
 /[O 行政院农委会林业试验所福山分所]/
 /[O 宋庆龄基金会]/
 /[O 上海轮胎橡胶（集团）股份有限公司]/
 /[O 中国驻日本大使馆]/
 /[O 美国白宫]/
 /前/[O 中国新华社香港分社]/社长/[P 许家屯]/
 /[O 清华大学计算机系人工智能实验室]/
 /[O 中保财产保险四川省分公司]/

5.2.3 国家（或国际）立法部门或行政部门标注为机构名

*/当选/[O 国会]/议员/
/[O 内阁]/改组/将会/在/[dat 八月底]/前/完成/
/前/[O 内阁官房]/长官/[P 山静六]/
/[P 刹瓦什]/向/[O 宪政法庭]/提出/动议/*

5.2.4 地名和机构名紧邻时的情况

地名和机构名的关系一般有以下两种情况：

- (1) 表示所属关系（如：‘法国航空航天局’，‘航空航天局’隶属‘法国’）。
- (2) 表示地理位置关系（如：‘北京邮电大学’表示大学位于‘北京’，而不是隶属于‘北京’）。

地名和机构名之间还可能有更复杂的情况，这里不予讨论。

5.2.4.1 规则一

如果机构名以一个地名开头，而且删除这个地名后所剩部分不再是一个具有特指性的机构名，那么该地名必须留在机构名中作为该机构名的一部分标注；

*/[O 北京大学]/
/[O 深圳中学]/
/[O 复旦大学专用集成电路与系统实验室]/
/[O 东南大学]/[O 深圳宝安设计院]/*

5.2.4.2 规则二

如果机构名前面还有一个或多个地名，那么该机构名与前面紧邻的地名应当分开标注。如：

*/[L 中国]/[O 北京大学]/
/[L 中国]/[L 广东]/[O 深圳中学]/
/[L 北京]/[L 昌平]/[O 十三陵抽水蓄能电站]/*

5.2.4.3 规则三

如果一个机构名的开头不是地名，那么当它前面邻接一个或多个地名时，只有其中与该机构名紧邻的那个地名需一起标注。例如：

*/[O 上海同济大学]/
/[L 中国]/[O 上海同济大学]/
/[O 湖北省武钢三中]/*

5.2.4.4 规则四

如果一个机构名本身以两个或两个以上并列的地名开头，则这些地名都要留在该机构名中。如果在它前面再出现其它地名时，一律同该机构名分开标注。但是如果上一级地名不能管辖下一级地名时，要把上一级地名标注在机构名内。

例如：

/[L 洛杉矶]/[O 亚太法律中心]/
 /[L 香港]/[O 中港贸易协会]/
 /[O 广东亚洲大酒店]/
 /[O 澳大利亚维多利亚投资公司上海办事处]/《》/
 /[O 澳大利亚维多利亚投资公司]/《》/

- 注：“广东”与“亚洲、澳大利亚与维多利亚”都不属于上、下级管辖关系，所以要把上一级地名标注在机构名内。

5.2.4.5 更复杂的情况

在更复杂的情况下，我们可能无法判定某机构名究竟是以一个还是两个地名开头的。这时可按规则 5.2.5 和 5.2.6 来处理。

例如，‘洛杉矶台北经济文化办事处’

究竟是 A： [L 洛杉矶]/[O 台北经济文化办事处]

还是 B： [O 洛杉矶台北经济文化办事处]

这时，默认的标注方式是 B (理由见 5.2.8)。

5.2.4.6 地名概念比较模糊的情况

如果该地名比较模糊，而标注者又没有足够的知识来判断某机构名的开头是否是一个地名。就标注到一个比较明确的地名，

例如：‘印度尼西亚莫巴蒂努山打腊航空公司’中的‘莫巴蒂努山打腊’不知道是不是地名。但至少知道一旦拿走了这个字符串，剩下的字符串已不构成专指性的地名。此时，按规则 2.5 的标注方式应是：

/[L 印度尼西亚]/[O 莫巴蒂努山打腊航空公司]/

/[O 河北沙岭子电厂]/

---‘沙岭子’是一个乡镇的地名，河北和内蒙古都有一个沙岭子镇，
 地名的概念比较模糊，故标注在机构名内。

/国际/著名/的/[O 加拿大 B+H 国际建筑师事务所]/

5.2.4.7 紧邻的地名和机构名不构成修饰关系的情况

一个地名后紧邻一个机构名，但它们不构成修饰关系，则一律分开标注。

/促进/了/[L 中国]/[O 东盟]/的/合作/

/在/[L 日内瓦]/[O 联合国]/&人/权&/会议/上/

更典型的例子需借助上下文来判断，如：

/促进/了/[L 中国]/[O 微软]/的合作/

/[O 中国微软]/即将/发布/新产品/

- 注：如果标注者不能判断它们是不是修饰关系，则默认为分开标注，如：

/[L 中国]/[O 微软]/

/*[O 美国众议院]*
 /*[L 重庆]*/*[O 长江救助打捞公司]*
 /*[L 日本]*/*[O 东京股市]* ----错误标注!
 /*[L 日本]*/*[L 东京]*/*股市* ----正确标注。
 /*[L 美国]*/*[L 华盛顿]*/*[O 三普证券公司]* ----错误标注!
 /*[L 美国]*/*[O 华盛顿三普证券公司]* ----正确标注。
 /*[L 华盛顿]*/*[O 美国国务院]*
 /*[L 瑞典]*/*[O 斯德哥尔摩国际和平研究所]*

5.2.5 会议、晚会、运动会等以‘会’结尾的短语是事件，不作机构名标注

/*泛*/*[L 美]*/*运动会*
 /*[L 中国]*/*[ord 第一届]*/*人工智能大会*
 /*[ord 第四届]*/*[L 中]*/*[L 法]*/*经济研讨会*
 /*[ord 第三届]*/*[L 海峡两岸]*/*水利科技交流研讨会*
 ----以上几例为事件，不是机构名。
 /*[O 中国人工智能协会]*
 /*[O 中国人工智能联合会]* ----为机构名。

当会议指议会(congress)或代表大会(chamber of deputies)时，应视为机构名。但是要注意：虽然议会或代表大会是机构名，但是议会或代表大会中的某一次会议是一个事件，不是机构名。为了更明确的区分各种情况，我们用以下例子说明：

/*通报*/*了*/*[O 八届政协]*/*[ord 五次]*/*会议*/*的*/*各*/*项*/*安排*
 /*[O 全国政协]*/*[ord 八届]*/*[ord 五次]*/*会议*/*将*/*于*
 /*听取*/*和*/*审议*/*[O 全国政协八届五次常务会议常务委员会]*/*报告*
 /*审议*/*[ord 八届]*/*[ord 五次]*/*会议*/*提案*/*审查*/*情况*/*的*/*报告*

• 注：‘八届五次’、‘五次’是一个事件，不应标注为机构名。但是这次会议的组委会、委员会应视为机构名。例如：

/*[O 八届全国人大]*/*[ord 五次]*/*会议*
 /*[O 政协九届一次会议]* --错误标注!
 /*[O 中国共产党第十五次全国代表大会]*
 /*[O 九届人大]*/*[ord 一次]*/*会议*
 /*[O 中国全国人大]*
 /*[O 中共十五大]*
 /*各级*/*人大*/*常委会* --不是专指，故不标。
 /*[O 中国科协]*/*[ord 第五次]*/*全国代表大会*
 /*[L 湖南省]*/*[ord 六届]*/*[ord 二次]*/*全委会议*
 /*向*/*同级*/*人民代表大会*/*或*/*人民代表大会常务委员会*/*提请*/*审议*

• 注：‘全国人民代表大会’和确指的省、市人民代表大会及其常委会、常务委员会需作为机构名标注。泛指的‘人大’、‘中央银行’、‘人民银行’、‘&农/发行&’不作为机构名标注。

/*[O 临澧县人大]*/*抓*/*村级*/*财务监督*/*一瞥* (*/**监督*/*广角*/*)*
 /*由于*/*各级*/*人大*/*代表*/*的*/*有效*/*监督*/*,*/*[dat 去年]*/*以来*/*该*/*县*/*各*/*村*/*村务*/*情况*/*出现*/*好转*/*,*/*/*

- 注：在地名‘国会大厦’中，‘国会’不可作为机构名标注，否则就出现嵌套了。
/L 国会大厦/
- 注：“联合国大会”及其简称“联大”都是词表词，但不可整体标为‘O’。如：
/O 联合国大会/于/dat 1992年/批准了/这一/条约/。
/P 沈国放/于/dat 27日/在/O 联大/全体/会议/上/表示/，
- 注：‘会’也可能出现在一般的机构名中，如：
/O 红十字协会/

5.2.6 用‘我们、我’等代词修饰的机构名，只对机构名进行标注

/我国/O 共产党/
/我们/O 清华大学/

- 注：根据上下文是确指的某公司、单位名称的简称要标注为机构，否则不标注！但如果在‘公司、集团’等词前面有‘本、我、该’等字样时，此处的‘公司、集团’不进行标注。其他特殊情况依据上下文进行标注。如：

/凡/《/O 克罗伏特缓冲器股份有限公司/股份/》/记名/的/持有/人/均/为/
本/公司/股东/。/

/我/公司/出/资/总/额/[mon 50 万元/

/O 港资陕西华懋实业公司/总/理/[P 商铭渔/，/受/O 公司/董事/会/委
托/来/到/O 咸阳市西北地勘局二一五医院/看/望/O 公司/保/安/员/[P
韩玉刚/，/

5.2.7 大使馆和领事馆的标注

- 当大使馆(或领事馆或其它外交使团)所代表的国家和所在地区相连时，整体标为机构名。如：
/后来/调/任/O 美国/驻/洪都拉斯/大使/馆/
- 当大使馆(或领事馆或其它外交使团)所代表的国家或所在地没有出现在上下文中，或者在描述范围内不连续，那么存在两种情况：
 - (1) 大使馆所代表的国家和大使馆（领事馆）相连，此地名和大使馆一起标记为机构名。如：
/前/往/[L 香港/]/的/[O 洪都拉斯/领/事/馆/
 - (2) 大使馆所在地和大使馆（领事馆）相连，此地名应单独标记，整体不作为机构名。如：
/[L 美国/]/在/通/过/驻/[L 金沙萨/]/大/使/馆/和/其/它/正/常/管/道/
- 注：虽然‘驻金沙萨大使馆’是一个连续的短语，但它的实际意思是‘美国（或X国）驻金沙萨大使馆’，而不是什么‘金沙萨（的）大使馆’。因此在这里‘大使馆’不视为机构名。

5.2.8 生产厂家要标注为机构名，产品则不标

这里定义的产品范围较广，不仅包括生产厂家生产出来的产品（如自行车等），还包括计算出来的产品（如：股票指数）、媒体产品（如：电视节目）

/[O 道琼]/工业/平均/指数/

---因为股票指数可以视为产品，那么‘道琼’就可以视为生产厂家。

/[O 纳斯达克]/指数/

---原因同前。

/[O 太原刚玉]/[dec 1 0. 5 8 1]/

/[O 咸阳偏转]/[dec 1 6. 1 1 2]/

/[O 深华发A]/[dec 1 5. 6 6 3]/

/[O 渝开发A]/

5.2.9 报纸、广播电台、电视台和杂志的名字要标为机构名

新闻媒体（如：报纸、广播电台、电视台和杂志等）的名字要标为‘O’；但报刊、电视栏目的名字不标。例如：

/[O 美国之音]/记者/表示/

/[O 人民日报]/海外/版/[ord 第三版]/

/《[O 泰晤士报]》/援引/一个/国际/专家/委员会/

/[O 中央电视台]/《/焦点/访谈/》/、/《/东方/时空/》/主持人/

/[O 武汉电视台]/《/科技/之/光/》/栏目的/《/科学家/，/您好/》/专栏/

/[O 美国《科学》杂志]/

/[O 美国探索电视网]/

/创办/《/[O 深圳房地产快讯]/》/

/办/好/《/[O 中外房地产导报]/》/

5.2.10 特殊情况

***民族不作为机构名

***泛指‘部队’不作为机构名

***政府不作为机构名

***学术或商务会议(conference, meeting)不作为机构名

***交易会不作为机构名

***运动会不作为机构名

***联赛不作为机构名

5.2.11 特殊情况示例

/[L 中国]/[L 天津]/出口/商品/交易会/

/[L 中国]/[O 天津出口商品交易会]/

----错误标注！

/[L 中国]/政府/

----不把‘政府’标为机构名。

/[L 非洲]/维持/和平/部队/

----不把‘部队’标为机构名。

/[L 中国]/公安/部门/

----不把‘部门’标为机构名。

/[O 中国公安部门]/

----错误的标注！

• 注：标注并列的机构名（O）时，连接词和标点符号不进入标注范围。例如：

/[O 上海]/、/[O 北京人类基因组研究中心]/
/[P 贺国中]/分别/[O 一]/、/[O 四]/、/[O 七团]/党代表

- 注：上述情况和标注并列的序数（ord）不同，连接词和标点符号是否进入标注范围取决于序数词所修饰的词语。例如：

/获得/个人/[ord 一、二、三等]/奖/
/[ord 一]/、/[ord 二]/、/[ord 三]/产业/
/书店/[ord 三、四层]/

- 注：‘中央’不作为机构名，但‘党中央’标为机构名。
/在/中央/的/领导/下/
/以/[P 胡锦涛]/同志/为/核心/的/[O 党中央]/周围/

5.2.12 地名和机构名容易混淆的情况

/[L 人民大会堂] ----地名。

/[O 五角大楼]/发言人/说/，

/[O 白宫] ----机构名。

/[O 克里姆林宫]/表示 ----机构名。

/在/[L 总统府]/分别/约见/了/多/位/[O 国民党]/中/常委/检察官/

- 注：‘总统府’标注为‘L’而不是‘O’。这是因为有的国家有多处总统府，所以不能把它们视为国家或政府的唯一代表。
- 注：下面的例子中出现的类似单位名称的，因不是确指，而且是出现在各种条令、合同中，适合任何一个‘省、市、县’的单位机构名称，所以不能作为一个机构名称标注为‘O’。如：

/本/合同/正本/[int 三份]/，/出租/人/、/承租/人/、/市/公证处/各/执/一/份/。
/副本/若干/份/，/报/市/经济/委员会/、/市/经济体制改革/委员会/、/市/财政
/局/、/劳动局/、/税务局/、/审计/局/、/工商/行政管理/局/、/[O 中国人民
银行]/市/分行/、/[O 中国工商银行]/、/市/分行/等/有关/部门/备案/
/本/合同/在/履行/中/如/发生/争议/，/双方/应/协商/解决/；/协商/不成/时/
/任何/一方/均可/向/工商/行政管理/局/合同/仲裁/委员会/申请/调解/或/仲裁/

第六章 数字串标注总则

数字串（Factoid）包括时间表达式（*TIMEX*）、数字表达式（*NUMEX*）、度量表达式（*MEASUREX*）和地址表达式（*ADDREX*）等4大类，27个小类，详见表1-1。标注数字串的一条重要原则就是：它的标记不得插入到词表词的内部（见1.5.2.4）。

6.1 时间表达式

时间表达式（*TIMEX*）包括日期（*dat*）、时间（*tim*）和时段（*dur*）三小类。所有小于一天的时间都被定义为时间（*tim*），如‘秒，分，小时’。一天或者大于一天的时间单位则属于日期（*dat*），如‘天，日，星期，礼拜，月，季度，年，五年，十年，世纪’等。时段（*dur*）通常也使用日期和时间中的单位，如‘月、年、时、分’等。对此标注者要注意区分。

将日期、时间同时段区分开来有时是困难的，下面分别给出它们的定义。

6.1.1 日期（*dat*）和时间（*tim*）的定义

日期和时间在一维的时间坐标轴上有相对确定的位置。小于一天的时间都被定义为时间。一天或者大于一天的时间则属于日期。

/tim 8点30分/

/dat 今天/[/tim 晚上/] ----‘晚上’是词表词。

/dat 昨天/[/tim 夜里/] ----‘昨天’和‘夜里’都是词表词。

/tim 昨夜/ ----‘昨夜、昨晚’都是词表词，只能整体标‘*tim*’。

/dat 昨/[/tim 晚/] ----错误的标注！

/dat 春节/ ----在每一年中，是比较固定一天或几天。

/dat 1999年/ ----以‘年’为单位，与别的年份相区别。

/在“[/dat 六五]/”/中/ ----以‘五年’为单位，与别的‘五年’相区别

- 注：严格地说，每一个‘*dat*’或‘*tim*’都占据了一个时间段，因此这里出现的‘*期间*’和‘*中*’，不能作为标注时段的理由。

/“[/dat 九五]/”/计划

/dat “九五”初/

/仅“[/dat 八五]/”/期间/就/达[/mon 一百一十五亿元/]。

/dat 下半年/ ----以‘半年’为单位，与‘上半年’相区别。

/dat 二十世纪/ ----以一百年为单位，与别的‘世纪’相区别。

/为/庆祝[/O 北京大学/]/建/校/[/dat 100周年/]/，/

/dat 民国八十六年/

/dat 民国六十年代/

/[/dat 八十八年下半年/]/及[/dat 八十九年/]/中央/统筹/分配/款/，/

/dat 公元二千年/

/*[dat 今年九月]*
 /“*[O 迈特兴华]*”/杯*[ord 首届/全国/象棋/大师/赛]*于*[dat 今日]*/收/秤
 /*[dat 1997年下半年]*/, /
 /可/于*[dat 农历年]*/前/迁居/。/
 /*[tim 第七十三分钟]*
 /*[tim 中午12点]*
 /*[tim 格林威治时间5时59分]* ----含有地名。
 /*[dat 第二天]*/*[tim 一大早]*, ----‘一大早’是词表词。
 /在*[dat 今年暑期]*/大学生/送/科技/下乡/活动/中/,
 /大约/*[tim 七点]*/到达 ----‘大约’不标。
 /*[tim 晚上大约七点]*/到达

- 注：‘大约’被两个‘tim’包围，分割不开，所以整体标上。这条标注规则遵照了 ER-99 和 MET-2 的标准。
- 注：事件‘戊戌变法、辛亥革命、甲午战争、五四运动’等都是词表词，其中的日期不标注。但当‘戊戌、辛亥、五四’单独出现时，应作为‘日期’来标注。例如：

/与/稍/后/的/辛亥革命/, /都/有/相通/的/地方/,
 /在/[L 香港]/回归/[dat 周年]/前夕/和/“/七七事变/“/纪念日/,
 /*[dat 戊戌]*/思潮/与/前此/的/洋务运动/,

6.1.2 时段（dur）的定义

时段既可以长于一天，也可以短于一天。它不同于日期和时间，在一维的时间坐标轴上没有确定的位置。例如：

/*[dur 三年]*
 /*[dur 半年]*
 /*[dur 四分之一世纪]*
 /*[dur 廿四个月]*
 /时间/长/达/*[dur 六分钟]*
 /*[dur 两个星期]*
 /*[dur 一个月]*/后/
 /曾/在/*[dur 5、6年]*/前/撰文/陈述/
 /早产/*[dur 十二周]*/左右/
 /大水/*[dur 十天]*/后/才/退/尽/
 /*[dur 一至两年]*
 /*[dur 一小时卅分钟]*
 /这/*[dur 几天]*
 /*[dur 卅天]*/会期/只/开/了/*[dur 九天]*
 /虽/经/*[dur 一整天]*/磋商/, ----‘一整天’不是词表词，但要标为‘dur’。
 与/洪水/奋战/*[dur 一天一夜]*/, ----‘一天一夜’也不是词表词。

时间表达式的标注细则详见第七章。

6.2 数字表达式

数字表达式（NUMEX）包括百分数（*per*）、钱款（*mon*）、频度（*fre*）、整数（*int*）、分数（*fra*）、小数（*dec*）、序数（*ord*）、比率（*rat*）等 8 小类。

6.2.1 百分数（*per*）

/*[per 百分之二十五]*/
 /*[per 百分之一.七]*/ ---虽然是小数，但要标作‘*per*’。
 /*[per 六点五百分点]*/
 /*[per 五成]*/以上/
 /*[per 六折]*/
 /*[fra 百万分之八]*/ ----注意标的是‘*fra*’而不是‘*per*’。
 /大约 /*[per 5%]*/ ----约数‘大约’不进入标注。

6.2.2 钱款（*mon*）

/*[mon 四亿元台币]*/
 /*[mon 43.6 亿美元]*/
 /*[mon 卅万元]*/
 /*[mon 四万五千块钱]*/
 /*[mon 四万五千元人民币]*/
 /只/增加/了/*[mon 几元钱]*/的/成本/
 /决不/乱/花/国家/的/*[mon 一分钱]*/。

- 注：同一笔钱的不同货币形式需分开标注。货币中的地名不标。
 /*[mon 26 万英镑]*/ (/ *[mon 43.6 亿美元]* /) /
- 注：‘约’是一个不确切的概念，故不标注。但‘上’、‘数’、‘好’要和数字串捆绑在一起标注。但‘近’作为特例，不与数词捆绑!!
 /约/*[mon 十万元]*/
 /大概/需要/花费/*[mon 上千万美元]*/的/投资/和/*[dur 3 年]*/左右/时间/，/
 /多/收入/*[mon 好几十元]*/

6.2.3 频度（*fre*）

/*[fre 数度]*/
 /*[fre 两次]*/
 /*[fre 26 次]*/
 /*[fre 十多次]*/
 /*[fre 多次]*/

- 注：动量词‘次’除了‘一次’不标注以外，其余的全部标注为‘*fre*’。
 /*[fre 一次次]*/

/[fre 再次]/
 /[fre 无数次]/
 /[fre 数次]/

6.2.4 整数（int）

‘int’标注的是数词和量词组合成数量词组。

/[int 卅七件]/
 /[int 一百卅项]/
 /[int 三种]/
 /[int 九个]/课室/
 /[int 几家]/
 /后/[int 几名]/
 /[int 十]/多人/
 /[int 四条]/断层/
 /[int 五十户]/
 /[int 百余名]/
 /[int 上万]/人潮/
 / “[int 双][int 百]”/方针/

- 注：“双百方针”是词表词，由于文中“双百”用引号括起，而且它们是两个数字，所以要分别按数字串标注。类似情况还有词表词“五四运动”，这是个事件不标。但是如果文中日期“五四”被引号括起，就要单独标为：/ “[dat 五四]”/运动/。又如“六一儿童节、六一国际儿童节、六一节”都是词表词。由于“六一”和“儿童节”是同一个日期，即使在文中“六一”被引号括起，也可以整体标为‘dat’，如：/[dat “六一”儿童节]/。
- 注：‘人次’应标注为‘mea’而不是‘int’，例如：
 /近[dur 3年]/中/，/该市/采取/多/形式的/农技/培训/近/[mea 万人次]/，
- 注：“数词+强”不一定表示序数，因此只单独标注数词为‘int’。例如：
 /在/这次/从/[int 十六]/强/到/冠/、/亚军/的/一次性/竞猜/中/，/
 /[O 宝钢]/为/跻身/世界/[int 500]/强/而/采取/的/重要/步骤/。

6.2.5 分数（fra）

/[fra 数倍]/
 /[fra 一半]/
 /[fra 千百倍]/
 /[fra 3/4]/
 /[fra 四分之三]/
 /[fra 百万分之三百六十四]/ ----注意标记是‘fra’而不是‘per’。
 /[fra 半个]/
 /[fra 4倍半]/ ----倍数是分数的一种表示，应标‘fra’。
 /[fra 4倍半]/

/[fra 4.5 倍]/ ----虽然‘4.5’是个小数，但不标‘dec’。

/有效/载/力/提高/[fra 2 至 3 倍]/

注：“过半数”是词表词，因此不作为分数‘fra’标注。例如：

/都/难以/获得/过半数/的/[int 2 0 7 张]/选票/，/

6.2.6 小数（dec）

/[dec 3.14]/

/[dec 三点一四]/

/看/了/那么/长/时间/的/电视/，/视力/依旧/[dec 1. 5]/

/我/有着/足以/令/我/自豪/的/[dec 1. 2]/视力/

----视力的多少是一个量级，没有单位，故按数量标注整数或小数。

/并/以/[dec 6139.69 点]/收盘/

/以/[dec 33.8]/收盘/

/比重/：/[dec 1.02]/

6.2.7 序数（ord）

/[ord 第一任]/

/[ord 第一期]/

/[ord 十六楼]/

/[ord 第三次]/世界/大战/

/[ord 首]/日/销售/欠佳/

/[ord 第二]/故乡/

/[ord 三等]/奖/

/[ord 前6 名]/

/地震/烈度/不/超过/[ord 8 度]/

/这/[ord 第二条]/尤为/重要/

/位居/金牌/榜/[ord 第二名]/

/[O 北京市]/[ord 首家]/就业/与/创业/组合/市场/

/[ord 1174 号]/文件/

/[ord 6 路]/汽车/

/[ord 六年级]/学生/

/[dat 今年]/读/[ord 大三]/

/发展/第一/产业/

----‘第一产业’是词表词。

/发展/[ord 第一]产业/

----错误的标注。

/阵风/[ord 五级]/

/[ord 一、二、三等]/奖/。

/他/亲手/接/治/[L 墨西哥]/[ord 首例]/艾滋病/患/儿/

6.2.8 比率 (rat)

/[rat 一比廿五]/
 /以/[rat 0 : 6]/失利/
 /上/一/届/世界杯/赛/就/以/[rat 1 : 0]/胜过。/
 /最终/以/[rat 三比三]/握手言和。/
 /用/原液/与/水/稀释/[rat 1 : 20 倍]/。/

数字表达式的标注细则详见第八章。

6.3 度量表达式

度量表达式 (MEASUREX) 包括年龄 (age)、温度 (tem)、角度 (ang)、长度 (len)、面积 (are)、容积 (cap)、重量 (wei)、速度 (spe)、加速度 (acc) 和其它 (mea) 等 10 小类。

6.3.1 年龄 (age)

/[age 卅五岁]/
 /[age 廿一岁]/
 /[age 六十五岁]/
 /[age 34 岁]/
 /[age 六十寿辰]/
 /[age 花甲]/老人/ -----‘花甲’是词表词。
 /如同/[age 年过半百]/的/老/妇/。/ -----‘年过半百’是词表词。
 /[P 李元]/、/[P 卞德培]/[int 两位]/先生/都/已/[age 年逾古稀]/。

6.3.2 温度 (tem)

/寒流/耍/酷/平地/[tem 6 °C]/
 /才/会/微/升/[tem 6.1 °C]/
 /但/平地/温度/还/会/下/探/[tem 5 °C]/左右/
 /积温/高/ ([tem 2800 度]/) / -----注意!
 /[tem 零下5 到 6 摄氏度]/

- 注：数字范围的标注方式详见 7.1.1。

/大约/[tem 5~7 °C]/
 /低温/反而/只有/[tem 10 °C]/~/[tem 12 °C]/
 /[tem 摄氏 19—24 度]/
 /[tem 摄氏 19 度]/-/[tem 24 度]/

6.3.3 角度（ang）

/钝角/就是/大于/[ang 90 度]/的/角/
/并/将/卫星/定点/在/[agn 东经 110.5 度]/[L 赤道]/上空。/
/震/中/位于/[ang 北纬 30.5 度]/， / ----详见 4.2.4.2

6.3.4 长度（len）

/开掘/到/[len 一米六七]/深度/时/
/高/[len 五米]/宽/[len 一百米]/
/[len 109×78 厘米]/
/[len 1 纳米]/=[len 十的负九次方米]/
/应用/于/紧/固/件/直径/为/[len 1 / 4 "]/（/[len 6m]/）/
/最高/速度/每秒/[len 360 米]/
/发生/每秒/速度/达/[len 四十二米]/的/大风。/
/[L 三峡]/截流/落差/在/[len 0.7-0.8 米]/之间， /

6.3.5 面积（are）

/[are 廿七公顷]/土地/
/占/地/[are 六百多公顷]/
/兴建/[are 五千坪]/大/的/厂房/
/[are 七百余坪]/
/每/套/住宅/面积/[are 140—160m²]/， /
/[are 997 万平方公里]/
/农田/[are 20 万亩]/

6.3.6 容积（cap）

/运输量/为/[cap 34 个立方]/
/[cap 一两箩]/谷子/
/选定/的/设计/流量/是/每秒/[cap 1.4 万至 1.9 万立方米]/。/
/工程/已/完成/土方/[cap 2 3 0 0 多万方]/， /
/全国/消费/了/[cap 2 5 万升]/啤酒。 /

6.3.7 重量（wei）

/[wei 九百至一千吨]/

/[wei 零点三公克]
/[wei 三千二百英吨]
/重/约/[wei 五、六公斤]/的/鲤鱼/
/[wei 十台斤]
/[wei 三点五公吨]
/产量/达到/[wei 数千万吨]
/[wei 几万吨]
/[wei 二十万吨]/级/以上/

- 注：ER-99 把上例标为：*[wei 二十万吨级]/以上/*。

6.3.8 速度 (spe)

/最高/速度/[spe 360 米每秒]
/打印/速度:[spe 12cps]

---- “cps” 表示“characters per second（每秒字符数）”。

6.3.9 加速度 (acc)

/抗震/能力:/地面/水平/加速度/≤/[acc 0.4m /s2]
/地面/垂直/加速度/≤/[acc 0.2m /s2]

6.3.10 其它度量表达式 (mea)

除了上面提到的度量单位元之外，物理、化学及其它度量单位的统一标注为‘mea’。

/额定/电压/至/[mea 660V]
/[mea 5.5 瓦特]
/参观/人数/达/[mea 620 万人次]
/工资/[mea 3500 元/人/月]
/[mea 25 元/公斤]
/风/压/不/超过/[mea 700Pa]/ (/相当于/风速/[spe 34m /s]/) /。
/迁移/到/[mea 千兆比特]/的/能力/能够/降低/拥有/总/成本/的/管理/方案/
/这些/快速/以太网/和/[mea 千兆位]/以太网/服务器/

6.4 地址表达式

地址表达式 (ADDREX) 包括电子邮箱 (ema)、电话 (pho)、传真 (fax)、电报挂号 (tel)、邮政编码 (pos) 和网址 (www) 等 6 种。

6.4.1 电子邮箱（ema）

/[ema exp@email.com.cn]/
/[ema cnhuang@msrchina.research.microsoft.com]/

6.4.2 电话（pho）

在标注电话号码时，要把国际区号、国内区号、本地区号等作为一个整体标注。如果有分机号码也要一并标注。当有多个分机号码时，要分别标注。如：

预约/订/位/电话/[pho 九五二八六二八]/
/洽/询/电话/[pho 二四九三一零二零]/
/订/席/专线/[pho (8610)-78906617]/
/查询/电话/是/(/[pho 零三八六二一一零零转二五二]/)/
/查询/电话/[pho 三六九九七二一转二三三一]/或/[pho 二三三二]/
/[pho 120]/
/[pho 119]/

6.4.3 传真（fax）

/全国/客户/服务/传真/：/[fax 010-58722727]/
/传真/号码/：/[fax 86-10-66665555]/
/公司/传真/：/[fax 86-10-66665555]/

6.4.4 电报挂号（tel）

/[O 搜狐公司]/电报挂号/是/：/[tel (8610)62726666]/
/电报挂号/：/[tel 86-10-66665555]/
/联系/电话/：/[tel 86-10-66665555]/

6.4.5 邮政编码（pos）

/[O 清华大学]/的/邮政编码/是/：/[pos 100080]/
/[L 安徽]/[L 阜阳]/地区/的/邮政编码/是/：/[pos 233600]/

6.4.6 网址（www）

/活动/报名/网址/：/[www http:www.acer.net/event/apply]/
/[O 蕃薯藤]/购物/网/（/[www http:shopping.yam.com/]）/

第七章 时间表达式标注细则

时间表达式（*TIMEX*）包括日期（*dat*）、时间（*tim*）和时段（*dur*）三小类。

7.1 日期（*dat*）

/[dat 明治三十九年]/ (*/[dat 公元一九零六年]/*) /
/[dat 大正十四年]/ (*/[dat 公元一九二五年]/*) /
/[dat 昭和二年]/ (*/[dat 公元一九二七年]/*) /
/[dat 清]/[dat 道光十四年]/
/[dat 清]/[dat 咸丰十一年]/
*/[dat 民国六十八年]/*拆除/后/迁到/[L 芦洲]/, */[dat 八十一年]/*间/又/扩建/
/[dat 一九九九]/
/[dat 一九九九年十二月三十号]/
/[dat 公元1990年4月22日]/
/[dat 旧石器时代]/
/[dat 八十年代]/
/[dat 下半年]/
/[dat 1989 财年]/ ----注意!
/[dat 1989 财年第三季度]/
/[dat 1990 上半财年]/
/[dat 1991 财政年度]/
*/[dat 秋季]/*报告/
/[dat 第四季度]/
/[dat 十五世纪]/
/努力/成为/一/名/高/素质/的/[dat 跨世纪]/人才。/
/值/此/[dat 世纪之交]/的/时候/,
/走/向/[dat 新世纪]/的/[L 中国]/律师/业/
/[dat 新旧世纪交替]/之际/
/黑色/[dat 星期一]/ ----注意!
/[L 北京]/在/[dat 23 号]/发表/了/报告/

- 注：数字串‘23 号’若不表示日期，则不标。
/[dat 五月上旬]/ ----‘上、中、下旬’要标注。
/科技/之/[dat 夏]/ ----注意!
/[dat 夏]/[dat 秋]/之间/
/[dur 一年]/中/四季/分明/ ----‘四季’是词表词不标注。
/[L 南极]/的/[dat 夏季]/
/[L 中国]/[dat 汉代]/
/[dat 春节]/ ----日期确定的节日要标注。
/[dat 肉孜节]/

- /[dat 开斋节]/
 /[dat 中秋]/时节/ ----注意‘时节’不标。
 /[L 美国]/的/[dat 独立日]/ ----美国独立日为每年 7 月 4 日。
 /[dat 27 年]/是/一个/多/事/的/年份/
 • 注：‘27 年’可能表示时段，标注者须根据上下文注意区分。
 /现在/是/[dat 26 号]/, /[dat 星期三]/
 ----同一个时间的不同表达，要分开标注。
 /现在/是/[dat 二月九号]/, /[dat 农历大年初三]/
 /大约/[dat 五月四日]/ ----‘大约，大致，大概’等词不标。
 /[dat 第二个十年]/
 /[dat 第二年 3 月]/
 /[dat 当年 9 月]/
 /[dat 今春]/ ----‘今春’不是词表词。

7.1.1 日期起讫表达式的标注

当日期表达式中有‘至、到’和连结符‘—’时，处在‘至、到’和连结符‘—’前后的日期表达式分别叫做‘前式’和‘后式’。如果前式和后式都是完整的日期表达式，则它们应分别进行‘dat’标注；否则前、后式要整体标注为‘dat’。

这条规则同样适合于其它各类数字串的标注，如：‘tim, dur, int, tem, wei, mon’等。其一般表达式为：

- / X+量词/到/X+量词/
 / X+量词/至/ X+量词/
 / X+量词/—/ X+量词/
 / X+至+X+量词/
 / X+到+X+量词/
 / X+—+X+量词/
 / X+、+X+量词/

例如：

- /[dat 三月三日]/至/[dat 三月卅一日]/
 /[dat 一月十八日]/到/[dat 廿一日]/
 /[dat 三月三至廿一日]/
 /[dat 二月十八日]/—/[dat 廿一日]/
 /于/[dat 今明两年]/陆续/推出/。/
 /[dat 民国五十五、五十六年]/
 /[dat 今明两天]/
 /[dat 今]/、/[dat 明]/[dur 两日]/
 /[dat 1980 年]到/[dat 1990 年 1 月]/

- 注：含有比喻意义的‘今天、昨天、明天、今日、昨日、明日’全不标注。
 /“一失足成千古恨，/同学/们，/看到/今天/的/我，/你们/是否/感悟/到/了/什么/?”/
 /尽管/炮火/已/消失/在/昨天/那/段/苦难，/
 /[O “四方”集团]/的/明天/将/会/更加/灿烂/美好。/

- 注：‘当年、同年、当月’等词语后有具体的日期时，要整体标注 **dat**，如果‘当年、当月、同年’等词语单独出现，而其前后有确指的日期时也要标注为 **dat**，否则不作标注！‘当日、当天’等词后有具体的时间时标注为 **dat**，否则不作标注！如：

*/[dat 当年 7 月]/在/[L 莫斯科]/举行/
/然后/于/[dat 同年 8 月]/奉调/回国。/
/[P 克林顿]/在/[dat 当月 1 3 日]/表示，/
/那是/[dat 当天]/[tim 中午 1 时]/的汇率/
/发言人/于/[dat 当日]/[tim 午夜]/发表/声明/*

7.1.2 ‘前、头、下’ + ‘时段 (dur)’ 应整体标注为 ‘dat’

*/[dat 头两个礼拜]/
/[dat 前 3 天]/
/[dat 今年头四个月]/
/比/[dat 上一年]/增长[per 1 0 . 4 %]/。/
/集中/研究/解决/[dat 下半年]/纠风/工作/如何/突出/重点，/
/[dur 两周]/前/
/[dat 1993 年之初]/ ----注意!
/[dat 公元之初]/*

7.1.3 当‘乾隆、康熙、道光’等表示年代时标注为 ‘dat’

当‘乾隆、康熙、道光’等表示年代时标注为‘dat’，而当‘乾隆、康熙、道光’等表示皇帝本人的名字时标为 **P**。如：

*/最近/发现/一/张/在/农家/珍藏/的/[dat 清代]/[P 康熙]/、/[P 雍正]/、/[P 乾隆]/、/[P 嘉庆]/、/[P 道光]/[int 五]/皇帝/诰封/圣旨/[int 九道]/，/
/收藏/了/自/[dat 清代]/[dat 乾隆]/年间/至今/各个/历史/时期/的/鼻烟壶/艺术/珍品，/*

7.1.4 朝代名的默认值为 ‘dat’

当朝代名被上下文确认为国家名时标注‘L’，否则默认为‘dat’。如：

*/如果/[P 刘伯温]/不是/一直/压抑着/对/[dat 元]/王朝/的/不满，/
/[dat 楚]/霸王/[P 项羽]/带领[int 两万]/兵/将，/
/只/带/[dur 三天]/粮食，/渡过/[L 漳河]/去/与/强大/的/[dat 秦]/兵/作战。
/结果，/[dat 楚]/军/大败/[dat 秦]/军。/
/[dat 吴]/王/[P 夫差]/战胜/了/[dat 越]/王/[P 勾践]/，
/[dat 战国]/时/[L 赵国]/良/相/[P 蔺相如]/曾/为/[L 赵国]/立下/汗马功劳；
/[P 唐雎]/出使/[L 秦国]/，/*

《[L 水浒]全传》/描述/的是/[dat 北宋末年]/震撼/[dat 宋]室/江山/的/[P 宋江]起义。/
/从/侧面/表现/了/[dat 清]政府/的/腐败/无能/, /激起/了/深/埋/在/人们/心底
/对/侵略者/的/敌视/和/对/[dat 清]政府/的/愤怒, /
/但是/, /战争/最终/因/[dat 清]政府/的/妥协、/投降/而/告/失败。/
/无奈/夜郎自大、/腐败/不堪/的/大/[L 清国]/武器/太/落后/, /

7.1.5 在“‘过去、今后、未来’+时段（dur）”等修饰成分不进入标注范围

/过去/[dur 3年]/中/,
/将/在/未来/[dur 几年]/内/出现
/未来/[dur 两天]/沿江/地区/仍有/中/到/大雨/, /
/[dat 今年七八月]/间/

7.1.6 词表词‘近年来、近些年、近几年来、近几年、几年来’等均不标注

按规定，词表词‘近年来、近几年、近几年、几年来、多年来、近些年’等内部的‘dat、tim、dur’都是不标的。但对非词表词则要分开标注。例如：

/[L 瑞士]/多年来/是/[ord 第一次]/。
/近几年/, / [L 中]/[L 菲]关系/
/近/[dur 五年]/来/
/时至今日/仍在/缓刑/期间。/ -----‘时至今日’是词表词。

7.2 时间

/[tim 凌晨零时]/
/[tim 清晨六时卅五分]/到/[tim 四十分]/
/[tim 凌晨二至四点]/
/[tim 中午十二时]/-/[tim 晚上九时]/
/[tim 上午十一时]/至/[tim 下午二时]/
/[tim 第七十三分钟]/
/[tim 格林威治时间5时59分]/ ----含有地名。
/[tim 下午当地时间5时59分]/
/[tim 九点整]/到达/[L 北京站]/
/[dat 九月十三日]/大约/[tim 七点]/到达/[L 北京]/

- 注：这里‘大约’不标。因为它虽被一个‘dat’和一个‘tim’包围，但是仍可以分割开。

7.3 时段

/[dur 两个星期]/
/[dur 一个月]/后/

/曾在/[dur 5、6 年]/前/撰文/陈述/
 /早产/[dur 十二周]/左右/
 /大水/[dur 十天]/后/才/退/尽/
 /[dur 一至两年]/
 /[dur 一小时卅分钟]/
 /这/[dur 几天]/
 /[dur 卅天]/会期/只/开/了/[dur 九天]/
 /[dur 10 个月]/
 /虽/经/[dur 一整天]/磋商/ ----‘一整天’不是词表词，但要标为‘dur’。
 /与/洪水/奋战/[dur 一天一夜]/，/----‘一天一夜’也不是词表词。
 /历经/[dur 一二十年]/创建/了/庞大/的/船队/，/
 /让/我们/全家/人/感动/了/[dur 好几天]/
 /在/水门/丑闻/[dur 四分之一世纪]/时/发表/的/评论。/

- 注：按照前面的原则：‘水门/丑闻/[dur 四分之一世纪]/时’在时间坐标轴上有比较固定的位置，因此应当标为‘dat’。但这种与事件（水门丑闻）相关的时间表达，在 ER-99 和 MET-2 中都是不标注的。这样，只有‘四分之一世纪’需要标注为‘dur’。

/[dur 十多年]/
 /[dur 几年]/以来/
 /在/[dur 半年]/时间/内/ ----注意：‘上半年’是 dat。
 /在/总结/[dur 14 年]/改革开放/经验/的/基础/上/

- 注：‘14 年’、‘30 年’也可能表示 dat。标注者要注意区分。

/我们/在/美国/奔波/了/[dur 30 年]/
 /[dur 27 年]/的/军旅/生涯/
 /整整/[dur 十五年]/ ----‘整整’不标。
 /大约/[dur 十年]/的/时间/ ----‘大约’不标。
 /[dur 十年]/来/
 /[dur 十几年]/的/时间/ ----注意！
 /[dur 十几年]/来/
 /[dur 十来年]/
 /[dur 数年]/
 /[dur 多年]/ ----ER99 不标。

7.3.1 ‘一年’都标为‘dur’

/新的/[dur 一年]/即将/开始/
 /硬/是/在/地下室/干/了/[dur 一年]/的/公司/
 /[dur 一年]/创/产值/效益/.../.../
 /聘金/为/[dur 一年]/[mon 900 万美元]/的/价码/

- 注：‘/这一年’、‘/那一年’中的‘一年’不是确指不作标注。

/这/一年/，/企业/增收节支/达/[mon 1 1 0 万元]/
 /在/[O 北大]/就读/的/那/一年/，/

- 注：‘整天、整日、整夜’一律标注为 dur，如：

/[dur 整天]/都/很/安静/，/

/还/东奔西走/[dur 整日]/忙/个/不停/, /
/让/人/[dur 整夜]/不得/入睡/

- 注：当‘年、月、日、周’等词修饰后面的‘工资、交易（销售）额、创汇’等词语时，要作为时段（dur）来标注。如：

/[dur 月]/收入/就/在/[mon 千元]/以上/
/[dur 年]/交易额/近/[mon 1 0 0 0 亿元]/。
/这/一/工程/[dur 日]/处理/污水/[cap 2 万立方米]/。

7.3.2 ‘一天’的标注有以下三种情况，需区别对待：

- 7.3.2.1 “前一天”，不论其前面有没有定语修饰统统标注为‘dat’（参见7.4.1）：

/[dat 前一天]/还/静止/的/电梯/[dat 今天]/动/起来/了/, /
/[L 香港]/[O 恒生]/指数/比/[dat 前一天]/下跌/[int 4 1 2点]/, /
/这次/[L 中]/[L 韩]/足球/对抗赛/是/在/[O 韩国队]/准备/赴/[L 法]/出征/
世界杯/的/[dat 前一天]/举行/的/, /

- 7.3.2.2 “一天”的意思是指时间段（24小时），标注为‘dur’：

/每人/每月/接待/来访/[dur 一天]/,
/[P 汤]/[P 尤]/杯/[dur 一天]/不/拿/回来/, /
/仅/[dat 5月31日]/[dur 一天]/, /[L 莫斯科市]/税/警/就/查出/[int 1 6
0 0 个]/违法/经营者/。/
/青年人/辛苦/忙碌/了/[dur 一天]/来/此/坐/坐/, /
/在/[L 墨西哥]/最后/[dur 一天]/的/访问/中/, /
/每/枚/多/赚/[mon 7 分钱]/, /[dur 一天]/下来/能/多/收入/[mon 好几十
元]/。

- 7.3.2.3 “一天”的意思相当于“有一天”，由于不是确指的日期所以什么也不标：

/但愿/有一天/我们/轻松地/说：/消费/着/是/美丽/的。/
/[dat 1 9 9 7 年]/的/一天/, /[P 吴佩民]/在/办公室/热情/接待/了/一个/素
不相识/的/中年/妇女/。/
/一天/[tim 下午]/, /记者/到/那/店/里/专门/拜访/了/[P 佛朗科]/师傅/。/
/一天/, /[P 列宁]/收到/一/封/前线/发/来/的/要求/支援/武器/和/服装/的/电
报/。/
/一天/上/晚/自习/回来/, /有/一/条/狗/总/跟着/她/, /
/一天/[tim 深夜]/, /一/人/酒后/拦截/过往/的/外地/车辆/, /
/一天/, /我/走过/他/的/门前/, /
/一天/[tim 晚上]/, /新上任/的/[L 河北省]/[O 栾城县委]/书记/[dat 六月八
日]/, /

- 注：“这/一天、那/一天”中的“一天”也非确指，所以也不标。

/记住/这/一天/, /也是/表达/我/对/[L 香港]/回归/祖国/的/预祝/。/
/[P 王龙雨]/从/上任/的/那/一天/起/, /

7.4 有关时间表达式的规则

7.4.1 ‘前(后)+日期|时间’要整体标注

/[dat 今年前五个月]/
/[dat 前三天]/

- 注：以下的标注是正确的：

/在/上半时/结束/前/[dur 1 分钟]/ ---‘上半时’是词表词。
/比赛/前/[dur 十分钟]/
/在/上/半场/[tim 第 2 7 分钟]/时/

7.4.2 反例——不应该标注的例子

‘刚才、最近、开始军备谈判以来、一会儿’等表示不确定时间的词语，不标。如果节日没有确定的时间，也不标。如：

/[L 印度]/国际/电影节/
/[L 中国]/旅游年/

7.4.3 特例

若两个短语属于不同的子类‘dat’和‘tim’，就需分开标注。

/[dat 2 月 12 日]/[tim 上午 8 点]/
/[dat 星期一]/[tim 8 点]/

- 注 1：时间中的地名，如‘北京时间下午 5 点’，在 ER-99 中不标注，而在 NET-2 中要标注。本规范按 NET-2 标注（参照前面的例子）。如果‘dat’和‘tim’分不开，就整体标注。

/[tim 北京时间 1997 年 2 月 9 号 19 点 28 分]/

- 注 2：‘去年、昨天、今早’等词在 MET-2 中要标，在 ER-99 中不标。本规范只参照 MET-2：

/[dat 去年上半年]/
/[dat 今年夏天]/
/[dat 今年三月一日]/
/[dat 去年春夏之交]/
/[dat 昨天]/[tim 夜里]/ ---‘夜里’是词表词。
/[dat 今天]/[tim 晚上]/ ---‘晚上’是词表词。
/[dat 今]/[tim 早六点]/ ---‘今早’不是词表词。
/[tim 早上六点]/ ---‘早上’是词表词。
/[dat 5 月份]/产品/出口/和/转口/总值/比/[dat 去年同月]/
下降/[per 3. 2 %]/,
/[dat 同一天]/[tim 晚上]/
/[dat 当日]/[tim 下午]/

- 注 3：‘当日’是词表词。如果在上下文中能确定‘当日、当天’或‘同一天’的具体日期时，就标注；否则不标。

/每日/[tim 上午1 1 时]/至/[tim 深夜3 时]/ ----‘深夜’是词表词。
 /[tim 昨夜]/ ----‘昨夜’是词表词。
 /每[dat 周四, 二, 一]/

- 注: MET-2 和 ER-99 对‘早上六点’的标注是相同的。但 ER-99 认为‘早上六点’与‘今早六点’不同。原因可以从英语的表达来理解: 前者是“6: 00 am”, 后者是“6: 00 this morning”。“this morning”在 ER-99 中被视为“相对时间”, 不标注。但在 MET-2 中, “相对时间”是要标的。本规范遵循 MET-2。

/[dat 11 月2 4 至2 7 日]/
 /[dat 3 月15 日]/至/[dat 17 日]/
 /[dat 1949 年]/-/[dat 1972 年]/
 /[L 美国]/南北战争/ ([dat 1 8 6 1—1 8 6 5 年]/) /中/
 /软件/最长/的/寿命/为/[dur 两到三年]/, /

---请注意这里日期范围的标注方式。

迄今 ----词表词不标-, MET-2 标‘今’。
 今后 ----词表词不标-, MET-2 标‘今’。
 晨练 ----词表词中的‘晨’不标。-
 晚宴 ----词表词中的‘晚’不标。-
 春联 ----词表词中的‘春’不标。
 他们的/今天/, /仿佛/就是/我们的/明天/。 ----泛指不标。
 参加/半决赛 ----‘半决赛’是词表词, ‘半’不标。
 双边/会谈 ----‘双边’是词表词, 因此‘双’不标。

7.4.4 ‘每年’和‘年’ 不标注

本规则也适用于‘月, 天, 小时’等其它时间单位。例如:

/年产值/.../.../
 /每年/创/产值/效益/.../.../
 /每年/收入/.../.../

第八章 数字表达式标注细则

数字表达式（NUMEX）包括百分数（*per*）、钱款（*mon*）、频度（*fre*）、整数（*int*）、分数（*fra*）、小数（*dec*）、序数（*ord*）、比率（*rat*）等 8 小类。以下是数字表达式的一些标注规则。

8.1 如果整数、分数、小数、序数后面有量词，数量短语要整体标注

例如：

/int 几千万盆/
/int 几家/工厂/
/一/家/ [int 5]/人/
/一/家/ [int 5 口]/人/
/铁人/int 三项/比赛/是/多/项目/的/综合/运动/，
/计算机/配置/：/586/以上/，/int 8 兆/内存/以上/
/打印/分辨率/：/mea 180dpi/

注：‘dpi’表示每英寸的点数，所以作为‘mea’标注。

/评为/“/[int 十]/星级/乡镇”/、/“/[int 十]/星级/支部”/

8.2 单纯的数字、词表词（包括俗语）中的数字都不作标注

如：

/自然数/5/和/6/都是/整数/
/大家/听/口令/，/齐步走/，/一/二/一/，/一/二/一/，/一/二/三/四/，/
/但是/卷子/上/的/“/6/”/还是/颠/巍巍/地/变成/了/“/8/”/。/
/[L 瑞士]/、/[L 西班牙]/、/[L 比利时]/、/[L 丹麦]/[int 四]/国/
/并/促进/了/[L 中][L 美]/两国/的/交流/与/合作/ ----‘两国’是词表词。
/并/促进/了/[L 中][L 美]/[int 两]国/的/交流/与/合作/， ----错误！
/垄断/了/[L 神奈川]/、/[L 青森]/等/[int 5]/县/的/交通/信号/维修/业务/。
/[L 两岸]/经济/合作/和/直接/三通/ ----‘三通’是词表词。
/[L 两岸]/经济/合作/和/直接/[int 三]通/ ----错误！
/到/[L 云]/[L 贵]/[L 川]/的/大三线/地区/， ----‘大三线’是词表词。
/到/[L 云]/[L 贵]/[L 川]/的/大[int 三]线/地区/， ----错误！
/十年寒窗/ ----词表词中的‘十年’不标。
/千载难逢/ ----词表词中的‘千载’不标。
/十年/九/旱/ ----非词表词。虚指的‘十年’不标。
/眼/观/六/路/，/耳/听/八/方/ ----非词表词。虚指的‘六、八’不标。
/利/在/千/秋/的/大事/ ----虚指的“千秋”不标。

- | | |
|---------------------|----------------------|
| /十年/如/一/日/ | ----虚指的‘十年’和‘一日’，不标。 |
| /万里[L 长城]/ | ----虚指的‘万里’，不标。 |
| /三皇五帝/ | ----‘三皇五帝’是词表词。 |
| /乌七八糟的/东西/几乎/扫荡/殆尽/ | ----‘乌七八糟’是词表词。 |
| /三大球/在/走/向/市场/时/ | ----‘三大球’是词表词。 |
| /第二次世界大战/的/反法西斯/斗争/ | ----‘第二次世界大战’是词表词。 |
| /三五成群/地/散落/着/警察/， | ----‘三五成群’是词表词。 |
- 注：‘一会儿，一起，唯一，付之一炬，一流，千方百计，一分为二，一切，二娃’等词表词中的数字一律不标。
 /本职/创/“/一流”/活动/
 /[int 亿万]/人民/
 /[int 百万]/民众/
 - 注：按照 ER-99，‘亿万、百万’不是一个抽象的数字，因此是要标注的。

8.3 ‘约、近’是一个不确切概念，故不同后面的数字串一起标注

‘上’、‘数’、‘几’、‘好’则要和数字串捆绑在一起标注，而‘约、近’作为特例，不与数字串捆绑。

- /大约/[int 12 亿]/人口/
 /约/[int 四五千]/人/在/[L 金边奥林匹克运动中心]/举行/集会/，/
 /约/[mon 十万元]/
 /近/[mon 千万元]/
 /大概/需要/花费/[mon 上千万美元]/的/投资/和/[dur 3 年]/左右/时间/，/
 /[O 省电力公司]/还/投资/[mon 好几百万元]/，/
 /多于/[mon \$90,000]/
 /[mon 几百万新元]/
 /统计/了/[int 上百种]/数字/，/
 /每年/都/要/花费/大量/外汇/引进/[int 上百套]/系统/
 /每年/搞/[int 一两个]/工程/，/
 /邀请/全国/近/[int 百名]/书法/名家/，/
 /近/[int 千名]/员工/
- 注：‘余、多’本不应标注，但当它们位于量词前分割不开，所以整体加以标注。
 /[mon 二十七万余元]/
 /[mon 五百多万元]/

8.4 钱款式中的地名

钱款表达式中的地名不论是单音节还是多音节的，Er-99 和 MET-2 都不标，否则就形成嵌套。

如果货币字符串在文本中单独出现，字符串中没有数字修饰，那么双音节的地名要标注为‘L’，单音节的地名不标注。例如非词表词‘泰铢’中的‘泰’不标。注意词表词‘日元’、‘美元’中的单音节的地名也不标。

/[mon 2000 新元]/
 /[mon 2000 新加坡元]/
 /泰/铢/汇率/稳定/在/[mon 3 8 铢]/—/[mon 3 9 铢]/兑/[mon 美元]/水平/
 /纷纷/抛/出/日元/购/进/[L 德国]/马克, /
 /[L 菲律宾]/比索/对/美元/汇率/也/下跌。/

8.5 钱款标注中的特例

MET-2 规定：如果没有表示钱款的单位，则不标。ER-99 则不然。本规范采用 ER-99 的规定。

/这/辆/汽车/值/[mon 20 万]/
 /卷标/上/的/价格/是/[mon 50]/
 /[O 纳斯达克]/跌/[int 140]/

8.6 频率的特例

/[fre 四年一度/ ----‘四年一度’并非词表词，但整体标注为‘fre’。
 /[fre 一年一度/
 ----‘一年一度’是词表词，整体标注为‘fre’。
 /主要/在/交流/[fre 50Hz]/, /额定/电压/至/[mea 660V]/
 ---交流电的频率是 50Hz (赫兹),即每秒变化 50 周。所以理应标成‘fre’
 而不是‘mea’。
 /频率/高/ ([fre 30—60KHz]/)
 /卫星/每年/发射/[fre 6 至 7 次]/。

- 注：‘又一次、再一次’全部标注为 fre，但‘/一次/又/一次’例外，不作标注。如：

/此间/舆论/[fre 又一次]/注意/到/[L 亚]/[L 非]/足球/的/差距/
 /精湛/演技, /[fre 再一次]/赢得/了/首都/观众/的/由衷/赞赏/

8.7 名词‘方’没有与之搭配的量词，因此可以和前面的数词直接结合

在‘我方’、‘校方’中的名词‘方’没有与之搭配的量词，因此可以和前面的数词直接结合，如：

/[int 三方]/已/就/[O 劳斯莱斯]/汽车/的/前景/达成/协定/, /

8.8 ‘一’相当于英语的冠词‘a’，一般不标

‘一’相当于英语的冠词‘a’，一般不标，但‘一倍’是例外，要标‘fra’。例如：

/一个/条件/
 /一/座/城市/
 /最大/的/企业/之一/

/荣立/一等功/ ----‘一等功’是词表词，不可标注。
/荣立/[ord 一等]功/ ----错误的标注！
/获/县/政府/新技术/推广/[ord 一、二等]/奖/。/
/我的/收入/是/她的/[fra 一倍]/ ----‘一倍’是要标的。

8.9 ‘一（1）+量词’不标注‘int’

8.9.1 ‘一+量词’是词表词的情况

词表词‘一个、一种、一类、一批、一次、一套、一阵’等作为数量短语不予切分，也不标注‘int’。其中有些量词重迭形式也是词表词，如‘一个个、一天天’，应保持其整词形式，而其它非词表词的数量短语和量词重迭形式都是要切开的。

/一个人/
/一个个/观众/
/一种/算法/
/一套/特种/邮票/
/一次/讨论/
/一/匹/黄骠马/
/一/栋/栋/楼房/
/一天天/暖和/起来/

8.9.2 词表词‘一起、一块、一道、一面’用作数量短语时应切开

词表词‘一起、一块、一道、一面’有副词和其它词性的用法，但当它们用作数量短语时一律切开，而且不标注‘int’。

/一/块/石头/
/一/起/交通/事故/
/一/面/镜子/

8.10 ‘一（1）+物理单位元’需按度量表达式标注

‘一（1）+物理单位元’（如米、公斤、摄氏度等）需按度量表达式（见 6.3）标注。如：

/[wei 一公斤]/大米/
/[mea 一度]/电/

8.11 分数词素‘半’

8.11.1 词表词中的词素‘半’不可标注为‘fra’（分数）

词表词如‘半价、半票、半饱、半身、半世、半辈子、上半时、下半场、半边’等，但不可把上述词表词中的词素‘半’标注为‘fra’（分数）。

/上/[fra 半]/场/比赛/[L 中国]队/未进/一/球/

/下半场/ ----词表词，是正确标注。

/下/[fra 半]场/ ----在词表词中插标‘fra’是错误的。

/目前/还/空闲/着/[fra 一大半]/的/营业/面积。/

/他们/之中/肯定/有/[fra 一多半]/人/没有/球/票/

/有/[fra 大半个]/篮球/场/那么/大/

- 注：当‘半’作为一个独立的词时要标注，标注的原则是：‘半+量词或名词’时标注，‘半+动词或形容词’时不作标注，如：

/下半场/后/[fra 半]/段/

/地处/偏僻/[fra 半]/山区/

/部分/企业/停产/或/半/停产/

/而/处于/半/死亡/或/休眠/状态，/

/干旱/半/干旱/地区/径流/造林/技术、/

8.11.2 以下的词表词不作为分数标注，而作为其它不同的数字串标注

/[dur 半年]/

/[dur 半天]/

/[tim|dur 半夜]/

/[int|age 半百]/

8.11.3 例外

半个西瓜中的‘半个’，与‘四半’中的‘半’概念不一样，前一个‘半’是指二分之一，后一个‘半’是量词，所以标注也不同。

/[fra 半个]/西瓜/

/[int 一个]/西瓜/分为/[int 四半]/

8.12 序数词素‘首’

8.12.1 词表词中的词素‘首’不可标注为‘ord’（序数）

词表中有许多词含有词素‘首’，如‘首创、首倡、首选、首发、首航、首飞、首演、首映、首战、首展、首席代表、首席科学家、首席执行官、首富、榜首、魁首、居首等。但不可把词表词中的词素‘首’单独作为‘ord’（序数）来标注。

/首席执行官/ ----正确标注。
/[ord 首席]执行官/ ----在词表词中插标‘ord’是错误的。

8.12.2 具有‘首+量词’结构的词表词或非词表词，应整体作为‘ord’标注

具有“首+量词”结构的词表词有：[ord 首届]，[ord 首次]，[ord 首批]，[ord 首位]，[ord 首例]等。

具有‘首+量词’结构的非词表词，如：

/L 北京市/[ord 首家]/就业/与/创业/组合/市场/
/P 满文军/则以/自己的/[ord 首张]/个人/专辑/
/dat 首日/销售/欠佳/

----这里‘首日’不能作序数词来标注，应标注为日期‘dat’。（详见 7.1）。

- 注：‘头版、头条’是词表词。它们和‘头一回’统统标注为‘ord’。如：
/在/[dat 4月11日]/的/《[O 人民日报]/》/[ord 头版]/[ord 头条]/社论/位置/发表/出来/，
/由于/是/[ord 头一回]/，/总/怕/有/个/闪失/，
- 注：“头”的上述标注不可类推到其它词组中，例如，
上/半场/表现/不好/，/头/[dur 10分钟]/甚至/有些/拖泥带水/。
----注：这里半场时间词表词，但不标注为fra。

8.13 ‘序数词+量词’结构，应整体作为‘ord’标注

/[ord 第一期]/
/[ord 第二]/故乡/
/[ord 三等]/奖/
/[dat 第一天]/ ----相对日期，标 dat，而不是‘[ord 第一]/天’。
/[dat 第二年]/ ----相对日期，标 dat，而不是‘[ord 第二]/年’。
/[O 波音]/747/ ----产品序号不标。
/地震烈度/不超过/[ord 8度]/
/这/[ord 第二条]/尤为/重要/，/
/位居/金牌/榜/[ord 第二名]/。/
/作为/大豆/行动/计划的/[ord 第二步]/
/[ord 1174号]/文件/

/[ord 6 路]/汽车/
 /[ord 六年级]/学生/
 /[dat 今年]/读/[ord 大三]/
 /发展/第一产业/ ----‘第一产业’是词表词。
 /发展/[ord 第一]产业/ ----错误的标注。
 /阵风/[ord 五级]/
 /通过/大学/英语/[ord 六级]/

- 注：联赛中的‘A/组、B/组’等不作为序数字串标注。如：
 /在/[L 里昂]/进行/的/世界杯/G/组/比赛/中/
- 注：“甲级、甲/A、乙/级、乙/A”等不作为序数‘ord’标注。如：
 /当即/停止/该场/比赛/主/裁判员/执法/全国/足球/甲/A/联赛/；
 /获得/[ord 前两名]/的/球队/晋级/甲/A/行列。/
 /[dat 1 9 9 8 年]/全国/[O 男篮]/甲/B/联赛/
 /判处/以/[P 东条英机]/为首/的/[int 7 名]/甲级/战犯/死刑。/

8.14 仅当形容词‘前’表示比赛名次时才和后面的序数结构一起标注

仅当形容词‘前’表示比赛名次，如‘前6名’、‘前四’（指‘前四名’）时，才和后面的序数结构一起标注为‘ord’。其余的情况如‘前两次’、‘前三组’、‘前三场’、‘前两项’等，‘前’都不得进入被标注的数字表达式。

/获得/[ord 前十名]/的是：
 /在/前/[int 两轮]/小组/赛/中/
 /列/前/[int 两位]/的是/[O 澳大利亚队]/和/[O 日本队]/。

8.15 文本中表示标号的数字不标

规范、条例中的条款标号，包括‘一、二、三、I、II、III、1, 2, 3、第一条、第二条、第三条’等，一律不予标注。只有当这些条款被正文引用时，才作为序号‘ord’被标注。例如：

/第二/, /制定/必要/的行规/, /行约/, /共同/规范/, /共同/遵守/, /
 /一/无/资金/, /二/无/场地/
 /一/靠/政策/调动/农民/的/积极性;/ /
 /二/靠/科技;/ /
 /一/是/继续/加强/农业;/ /
 /二/是/采取/措施/稳定/物价/, /抑制/通货膨胀;/ /
 /1/. /自卑/的/羞耻/感。/
 /2/. /依赖/的/恐惧/感。/
 /（1）/加强/爱国主义/的/宣传/教育。/
 /（2）/加强/正确/的理想、/信念、/人生观、/价值观/的/宣传/教育/。

“第+数词+条”视为词表词，但作为文中陈述的标号时不标注‘ord’。仅当其在文中被引用时才作为‘ord’标注。例如：

/第一条/, /消费者/永远/是/对/的/;

/第二条/、/如果/消费者/真的/错/了/，/请/参照/[ord 第一条]/。/

- 注：当上述数字表示等级序号时，则要标注为‘ord’。例如：

污秽/等级：/[ord I]/、/[ord II]/、/[ord III]/、/[ord IV]/。

8.16 人名、地名、机构名中的数字，不单独标注‘int’

*/P 佐腾一郎/
/L 梅竹蹊六十七号茶花庄/
/O 子弟一中/
/O 三明市/
/任/队长/的/[O 1205 钻井队]/*

8.17 外文字符串的标注

由于外文的词与词之间都有空格作为分隔符，因此无需再去切分，只在标点符号的前后加切分标记。遇到字母词、名称缩写等情况也不作切分，如：‘/COM/经济/’（网络经济）、‘/E/产品/’（电子产品）、‘/卡拉/OK/’等。

*/Good morning/、/everyone/、/
/最近/引进/一/台/JT-ESWL-III/型/体/外/震波/粉碎/肾结石/机/、/
“/[L ZHONG HUA REN MIN GONG HE GUO]/”、/这是/[L 中华人民共和
国]/的/汉语拼音/。/
“/Brother/、/I love you all the time/、/Thank you very much/! /”
“/Happy birthday to you/! /”
/Dip one end of a straw in the solution/./Blow gently through the straw/./
/A soap bubble forms/./What happens when you keep on blowing?/
/The bubble bursts because the pressure inside the bubble is more than the
pressure outside the bubble/./*

8.18 数学公式和机型标号均作为一个整体来切分和标注

例如：

*/ΔS=/[len 12 (S1+S2) mm]/
/IEC298. 265. 129. 694. 420. 56. 529. 932/
/GB3804. 3906. 11022/
/IEC60129A2/ (/ [dat 1996]/)
UES-K3 /2/
/UEMC40K8U /1/
/1V/FJ220001R2/
/SFL12 /17.5/IVDP575303RI/
/SFL24A/IVDP5753/O2RI/*

第九章 分词歧义消解细则

本章中的歧义切分实例是从微软亚洲研究院 2 3 7 万词训练语料、10 万词测试语料和 20 万词散页语料中抽取出来的。这些歧义字段可粗分为交集型歧义(OAS)和组合型歧义(CAS)两大类。交集型歧义又包含用正反向最大匹配(MM)算法侦查不到的所谓隐藏的CAS。下面就分别介绍不同歧义字段的消解规则。

9.1 交集型歧义字段(OAS)

9.1.1 交集型歧义字段示例

由于交集型歧义字段的例子太多，不便穷举，所以下面只列举少量实例供参考。

- (1) /矛头/所/指/正是/以/包/代/管/、/负/盈/不/负/亏/、/
- (2) /[L 四川]/一/私营企业/家/向/下岗/女工/捐款/
- (3) /柚木/购/进/后/市场价格/大跌/、/
- (4) /图/为/[O 保险公司]/向/受灾/企业/赔/付/现场/
- (5) /地方政府/亟需/在/加强/压/锭/监管/力度/方面/下功夫/、/
- (6) /与/厂/内/存留/的/旧/纱/机/一并/销毁/。/
- (7) 经/请示/、/自行/将/本/厂/经/改造/的/应/压缩/设备
- (8) 擅自/新增/棉纺/生产能力/、
- (9) /有人/钻/政策/空子/、/骗/财政补贴/。
- (10) 日益/猖獗/的/走私/犯罪/活动/、
- (11) /他/在/教务/活动/中/积极/研究/、
- (12) 对/全/山/的/商业/网点/和/摊/区/重新/进行/了/规划/和/建设/。
- (13) /加强/各级/领导班子/建设/。
- (14) 全体/员工/开展/了/“/人家/学/我们/、/我们/怎么办/”的/大/讨论/、
- (15) /[O 欧佩克]/提高/原油/配额/和/暖冬/等/因素/影响/、
- (16) /保护/国家/和/人民群众/的/生命/财产/安全/。
- (17) /以/维护/民族团结/为/己任/、
- (18) /各级/领导干部/要/站/在/党/和/国家/全局/的/高度/、
- (19) /只有/坚持/解放/思想/、/实事求是/的/思想/路线/、
- (20) /呈现/了/“/部队/添/战斗力/、/企业/增/生产力/、
- (21) /共建/双方/通过/自上而下/层/层/签约/、
- (22) /电力/部门/还/专门/建立/了/正规/的/转业军人/业务/培训/机制/、
- (23) /这/条/线/不/停电/、/官兵/跳伞/太/危险/了/。
- (24) 如同/[L 华中]/电网/强大/的/发电机/群/按照/同一/频率/转动/一样/、
- (25) 通过/举办/一些/全/集团/参与/的/拥军/活动/、

- (26) /在全国/工业/[ord 5 0 0]/强/中/名列前茅/的/大型/企业集团/。
 (27) /本/次/检测/中/性能/系数/最高/者/。
 (28) /不是/主张/所有/的/会议/都/开/成/电视电话/会议/。

9.1.2 隐藏的交集型歧义字段

隐藏的交集型歧义字段是指那些用正、反向最大匹配（MM）算法无法侦查到的交集型歧义字段。

注：以下例句中，双百分号‘%%’右面为改正后的切分。

- (1) /[L 新疆]/经济/社会/发展/一定/会展/现出/越来越/美好/的/前景/
 %%/[L 新疆]/经济/社会/发展/一定/会/展现/出/越来越/美好/的/前景/
 (2) /成立/了/专/司空/中和/地面/服务/质量/监管/的/服务/质量/督察/办公室/, /
 %%/成立/了/专/司/空中/和/地面/服务/质量/监管/的/服务/质量/督察/办公室/, /
 (4) /其/内/容/或/规则/已/译/成/[int 1 5]/国/语言/, /
 %%/其/内/容/或/规则/已/译/成/[int 1 5]/国/语言/, /
 (5) /这/一/发/现有/可能/加速/艾滋病/新药/和/疫苗/的/研制/。 /
 %%这/一/发/现/有/可能/加速/艾滋病/新药/和/疫苗/的/研制/。 /
 (6) /[L 韩国]/对/日出/口中/, /
 %%/[L 韩国]/对/[L 日]/出口/中/, /
 (7) /恰/在/此时/, /奉/党委/派赴/[O 共产国际]/工作/的/[P 张太雷]/于/[dat 8 月]/回国/
 %% /恰/在/此时/, /奉/党/委派/赴/[O 共产国际]/工作/的/[P 张太雷]/于/[dat 8 月]/回国/
 (8) /站/在/建/设有/[L 中国]/特色/社会主义/全局/
 %%/站/在/建设/有/[L 中国]/特色/社会主义/全局/
 (9) /金融/危机/就/可能/会演/变为/经济危机/
 %%金融/危机/就/可能/会/演变/为/经济危机/
 (10) /如/少数/司机/在/东侧/门楼/外/道路/上/违章/占/道/停车/, /
 %%如/少数/司机/在/东侧/门楼/外/道路/上/违章/占/道/停车/, /
 (11) /有/关/部门/要/下/决心/下/力/气/管/好/电子/游戏/室/。 /
 %%有/关/部门/要/下/决心/下/力/气/管/好/电子/游戏/室/。 /
 (12) /相/近/似/的/设施/化/保护/菜地/面积/达/[are 1 3 0 0 万亩]; /
 %%/相/近/似/的/设施/化/保护/菜地/面积/达/[are 1 3 0 0 万亩]; /
 (13) /表/明/了/财/政/部/门/对/落/实/科/教/兴/国/战/略/采/取/的/实/际/行/动/。 /
 %%/表/明/了/财/政/部/门/对/落/实/科/教/兴/国/战/略/采/取/的/实/际/行/动/。 /
 (14) 儿/时/站/在/家/门/口/向/四/面/望/, /
 %%儿/时/站/在/家/门/口/向/四/面/望/, /
 (15) /就/不/可/能/正/确/地/理/解/和/执/行/党/的/路/线/方/针/政/策/, /
 %%/就/不/可/能/正/确/地/理/解/和/执/行/党/的/路/线/方/针/政/策/, /
 (16) /使/得/高/校/中/原/有/的/个/别/的/知/识/物/化/行/为/迅/速/扩/展/为/一/种/专/门/职/能/。 /
 %% /使/得/高/校/中/原/有/的/个/别/的/知/识/物/化/行/为/迅/速/扩/展/为/一/种/专/门/职/能/。 /
 (17) /这/是/大/都/市/里/的/一/个/皮/货/修/理/店/, /
 %%/这/是/大/都/市/里/的/一/个/皮/货/修/理/店/, /
 (18) /需/招/集/体/制/女/性/业/务/员/[int 四名], /
 %%/需/招/集/体/制/女/性/业/务/员/[int 四名], /

- (19) /连同/应/交/费用/以/划/支票/寄/还/。 /
 %%/连同/应/交/费用/以/划/支票/寄/还/。 /
- (20) /我们/一口气/跑/到/家/门口/的/一/棵/大树/前/， /
 %%/我们/一口气/跑/到/家/门口/的/一/棵/大树/前/， /
- (21) /“唉/， /要是/好/好/复习/， /可/不/会/考/得/这样/糟/。” /
 %%“唉/， /要是/好/好/复习/， /可/不/会/考/得/这样/糟/。” /
- (22) /特/制/定/本/规/定/。 /
 %%/特/制/定/本/规/定/。 /
- (23) /股/东/会/所/议/事/项/作/成/会/议/记/录/， /
 %%/股/东/会/所/议/事/项/作/成/会/议/记/录/， /

9.2 组合型歧义字段 (CAS)

组合型歧义字段在真实文本中大量出现，有的是比较常见的，有的是非常罕见的。尤其是有的 CAS 即使根据上下文也很难判断其正确切分，如‘正在、就是、还是、只有、只是、一道、一起’等等。因此有必要针对那些高频的 CAS 逐条加以说明。

9.2.1 常见的组合型歧义字段

下面对一些常见的组合型歧义字段加以解释。

9.2.1.1 数词‘一’和量词组成的 CAS

词表词‘一个、一种、一类、一批、一次、一套、一阵’等作为数量短语不予切分，也不标注‘int’。其中有些量词重迭形式也是词表词，如‘一个个、一天天’，应保持其整词形式，而其它非词表词的数量短语和量词重迭形式都是要切开的。（详见 8.9）

/一个/人/
 /一个个/观众/
 /一天天/暖和/起来/
 /一套/特种/邮票/
 /一次/讨论/
 /一/匹/黄骠马/
 /一/栋/栋/楼房/

词表词‘一起、一道、一样、一手、一面、一口、一头、一气’等既可以用作‘连词、副词、名词或形容词’等，又可以切分开来成为数量短语。但像‘一套’这样的词表词，除了数量短语的用法以外，不再有其它用法，因此不存在切分问题。词表词‘有一套’是‘有本事’的意思时，也不切分。这类词的切分问题只能逐个加以描述。

9.2.1.2 动量词‘次’与频率‘int’的标注

动量词中只有‘次’被标注为频率‘fre’，如‘[fre 再次]、[fre 数次]、[fre 一次次]、[fre 无数次]、[fre 好几次]’，而‘遍、回、趟’不标注为频率，‘一/遍、一/回、一/趟、一次（词表词不切分）、一/遍/又/一/遍、一/回/又/一/回、一/趟/又/一/趟，一次/又/一次’也不标注为‘fre’。这条规则的理由如下：

(1) ‘遍’表达的是动作从开始到结束的全过程；‘次、回’描写动作的重复；‘趟’只用于表示行走意义的动词。‘去一趟’可以说成‘去一次’、‘去一回’，但‘做一次’、‘做一回’不能说成‘做一趟’。

‘遍、次、回’有时可通用，如‘你再唱一遍’，可以说成‘你再唱一次’或‘你再唱一回’而意思不变。但单纯表示动作数量时，只用‘次’，不用‘遍’，如‘他表示了多次’、‘敌人的三次进攻都被击退了’。

‘次’与‘回’区别在于，‘次’既用于书面语又用于口语；‘回’只用于口语。如‘多次、数次’等带文言色彩的短语，就不能说成‘多回、数回’。

(2) ‘这本书我看了一遍’，是指从书的开头到末尾的全过程。‘这本书我看了一次’，着重指看的次数，不指看的全过程。

9.2.1.3 ‘一(1)+物理单位元量词’构成度量表达式

当‘一(1)’后面是长度、重量等物理单位元时应分别按度量表达式标注为‘len,wei’，如‘[len 一米]、[wei 1 公斤]’（见 8.10）。

9.2.2 CAS 示例

下面是一些常见 CAS 的切分规则和示例。

(1) ‘人为’作形容词时不切分。

(1a) 而是/深究/灾难/中/的/人为/因素/。

(1b) 以/人/为/本/

(2) ‘为人’：

(2a) 也/包括/最/基本/的/为人/处事/的/行为/准则/

(2b) /始终/主宰/着/他/的/为人/之/道/和/为/艺/之/方/。

(2c) /我/把/不大/为/人/所/知/的/一些/往事/写/下来/，

(3) ‘一起’：作名词和副词使用时不切分，作为“数+量词”时切分。

(3a) /和/市民/一起/聊天/，/听取/群众/反映/。

(3b) /[dat 4 月 1 7 日]/发生/在/[L 北京]/[L 海淀区]/[L 阜石路]/的/一/起/车祸/，/

(4) ‘一点’：形容词，意思是‘少许’，不切分。但作为数量短语时要切开。

(4a) /文/中/还有/一点/小/误差/，/也/顺便/提/提/。/

(4b) /都/清楚/地/意识到/了/这/一/点/，/

(5) ‘一道’：作副词使用时不切分；作为“数+量词”时切分。

(5a) /而且/要求/未来/的/丈夫/同/她/一道/挑起/照顾/[P 穆]/大爷/的/担子/。

(5b) /已/成为/百里/油田/的/一/道/风景/线/。

(5c) /在/我/的/前额/刻下/了/一/道/道/弯曲/的/青春/印记/。

(6) ‘一面’：作名词和副词使用时不切分，作为“数+量词”时切分。

- (6a) /虽然/在/现代汉语/里/含有/贬义/, /但/其/积极/的/一面/应该/肯定/。
- (6b) /一面/学习/, /一面/实践/, /贯彻/到/筹组/[L 澳门特别行政区]/的/工作/中/去/。
- (6c) /爱/是/一/面/辽阔/光滑/的/回音壁/, /微小/的/爱意/反复/回响/着/，
- (7) ‘一口’：作形容词和副词使用时不切分，作为“数+量词”时切分。
- (7a) /P 崔]/又/一口/回绝/并/与/其/发生/争吵/。
- (7b) 要不/则/是/一/脸/匪/相/或者/一口/痞/气/，
- (7c) 不由得/倒/吸/了/一/口/冷气/打/了/一个/寒战/，
- (7d) 一/口/大/锅/解决/了/[int 两家]/的/再就业/难题/。
- (8) ‘一手’既有名词和副词的用时，又有“数+量词”的用法，但在文本中一律不予区分。
- (8a) 所有/的/为/官/为/政/者/都/能/写/一手/好/文章/，
- (8b) /整个/事件/是/他/一手/策划/的/，
- (8c) /他/一手/划水/， /一手/搂/着/女/青年/游/向/岸边/。
- (9) ‘一头’：作名词和副词使用时不切分，作为“数+量词”时切分。
- (9a) /二来/街道/一头/联/着/片/内/的/企业单位/， /一头/联/着/居民/，
- (9b) /此后/， /他/一头/钻进/常年/云雾/缭绕/的/云雾山/， /拜访/民间/郎中/，
- (9c) /一/头/经过/救助/已/恢复/健康/的/灰/鲸/ 『/[P 杰杰]/』 /
- (10) ‘一路’：作名词和副词使用时不切分；作为“数+量词”时切分。
- (10a) 有时/公共汽车/挤/不/上/， /干脆/快步/当/车/一路/小跑/。
- (10b) 我们/一路/攀登/来到/[P 王永祥]/简陋/的/护林/小屋/。
- (10c) 另/一/路/是/探索/[L 火星]/、/[L 木星]/等/星球/。
- (11) ‘一下’：用作副词和数量词使用时不切分；当‘一’作副词‘下’作动词时要切分。
- (11a) /只要/通融/一下/， /既/能/得到/一/笔/大钱/， /又/能/保持/友情/。
- (11b) /相互/拍打/一下/: /“你/猜/[rat 几比几]/?” /
- (11c) /书包/斜/背/在/肩/上/， /带子/太/长/， /随着/步子/一/上/一/下/跳跃/着/拍打/在/屁股/上/。
- (12) ‘一片’：作形容词使用时不切分；作数量短语时切分。
- (12a) /台上/台/下/那/一片/亲切/和谐/的/气氛/，
- (12b) /融入/一片/[dat 夏日]/的/浓绿/之中
- (12c) /地板/上/看/不/到/一/片/碎/纸屑/。
- (12d) /宽宽大大/的/粽/叶/， /她/总/要/一/片/片/洗净/。
- (13) ‘一则’：作副词使用时不切分；作数量短语时切分。
- (13a) /一则/表达/对/同乡/画/马/大师/[P 徐悲鸿]/的/敬仰/， /二/则/愿/家乡/建设/如/骏马/奔腾/一日千里/。
- (13b) /[L 法国]/报纸/刊/出/一/则/特写/，
- (14) ‘不见’：是动词‘见’的否定形式不切分。当它同前面的动词形成‘V/得/见、V/不/见’的可能式动补结构时，要切分。类似的可能式动补结构还有‘V/得/下去、V/不/下去’，

‘V/得/来/、V/不来/’，‘V/得/起/、V/不起/’，‘V/得/了/、V/不了/’，‘V/得/成/、V/不成/’，‘长/得/大/、长/不大/’等。

(14a) /全/都是/“/不见/兔子/不/撒/鹰/”/。

(14b) /人/武部/就/看/不/见/一/盏/长明灯/，

(15) ‘不对’：作形容词表示‘不正确’时不切分；如果‘对’作为介词，就要切开。

(15a) 父母/这么/想/当然/不对/， /可/也/不能/全/怪/他们/的/愚钝/和/落后/。

(15b) /中国/主张/和平/的/外交/政策/， /中国/不/对/任何/国家/构成/威胁/。

(16) ‘不等’：作形容词表示‘不相同’时不切分；如果‘等’作为动词，就要切开。

(16a) /按照/用户/要求/生产/大小/不等/的/编织/塑料/袋/，

(16b) 不/等/妻子/说/什么/， /他/自己/悄悄/地/找/开/了/出路/。

(16c) /时间/不/等/人/

(17) ‘不下’：表示‘不少于’时不切；作为动词‘下’的否定式和可能式动补结构（见14），就要切开。

(17a) /每天/她/经手/的/业务/不下/[int 百笔]/，

(17b) /架子/还/放/不/下/， /面子/还/丢/不/开/，

(17c) /刑/不/上/大夫/， /礼/不/下/庶人/

(17d) /[L 俄罗斯]/整个/国家/开支/居/高/不/下/，

(18) ‘不成’：作动词、形容词和助词使用时不切分；当它作为可能式动补结构（见14）时，一律切开。

(18a) /难道/自己/这/一辈子/就/这么/过/不成/？

(18b) /毛虾/已/不成/汛/，

(18c) 攀/“/亲/”/不成/反/折本/，

(18d) /往往/是/有/点/而/形/不/成/网/，

(19) ‘上下’：用作动词时一律切开，如“上/下/火车”；用作名词（包括并列意义）时则不切，如“上下/两册”。

(19a) /经过/上下/的/共同努力/，

(19b) /上/下/车/、/船/， /须/待/车/、/船/停/稳/后/先/下/客/后/上/客/，

(20) ‘从前’：作时间名词时不切分；如果‘从’作介词‘前’作方位词，就要切开。

(20a) /有的/是/从前/在/队/中/当/板凳/球员/，

(20b) /导致/美元/对/马克/的/汇价/从/前/一/交易/日/的/[rat 1 比 1·7 7 6 6]/降至/[rat 1 比 1·7 7 6 2]/。

(20c) /从/前/不久/[L 深圳]/一家/公司/大规模/地/恶意/抢/注/商标/案/，

(21) ‘以为’：作动词时不切分。

(21a) /以为/强大/的/[P 卡斯珀罗夫]/恢复/了/他/的/本来面目/。

(21b) 有/一些/干部/想/不/通/， /以为/是/搞/形式/， /出风头/。

(21c) /我们/引/以/为/自豪/的/风格/多少/应/有些/改变/了/。

(21d) /代表/们/以/为/人民/高度/负责/的/精神/， /提出/批评/和/意见/。

(22) ‘正当’：作形容词时不切分。

(22a) /我们/是否/能/以/某/种/不/正当/的/方式/反对/，

(22b) /正/当/禾苗/生长/关键/时期/， /

(23) ‘正在’：

(23a) /对/各地/已/建成/尚未/售出/和/正在/建设/的/住房/，

(23b) /记者/正/在/回/[L 巴黎]/的/高速/列车/上/。

(23c) /[O 世界卫生组织]/正/在/[L 科特迪瓦]/召开/国际/会议/，

(24) ‘会上’：

(24a) /[L 苏州]/等/省市/及/有关单位/在/会上/介绍/了/经验/，

(24b) /在/[O 国际泳联]/[dat 二十四日]/举行/的/听证/会/上/，

(25) ‘台上’：

(25a) /表演/完/节目/后/竟/在/台上/掩/面/而/泣/。 /

(25b) /预赛/是/在/有/围/绳/的/拳击/台/上/

(26) ‘走向’：用作名词时不切分；用作‘动词+介词’时，一律切分。

(26a) /[L 北京]/输/气/管道/工程/线路/走向/示意图/（/示意图/：/[P 孙伟]/绘/） /

(26b) /迈出/了/我国/航天/事业/走/向/世界/的/[ord 第一步]/。

(27) ‘才能’：用作名词时不切分；用作‘副词+能愿动词’时，必须切开。

(27a) /但/如果/施展/才能/的/空间/很/大/， /而且/能/充分/发挥/所/学/专长/， /不妨/一/试/。

(27b) /勤奋/才/能/有/真知灼见/；

(28) ‘人才’：用作名词时不切分。

(28a) 要/想/成为/[dat 跨世纪]/人才/， /光/有/专业知识/不够/，

(28b) /这/恐怕/只有/浪漫/的/[L 法国]/人/才/想/得/出来/。

(29) ‘上来’：作趋向动词和动词时不切分；当‘上’作方位词时，就要切开。

(29a) /一/届/新/班子/上来/以后/， /

(29b) /把/工作/重点/转移/到/社会主义/现代化/建设/上/来/， /

(30) ‘上去’：作趋向动词和动词时不切分；当‘上’作方位词时，就要切开。

(30a) /显然/是/上/了/学/的/[L 瑶族]/娃子/写/上去/的/。 /

(30b) /把/科研/技术/成果/转移/到/社会/应用/上/去/。 /

(31) ‘上前’：

(31a) /他/的/四个/弟兄/挨次/伸出/手/来/上前/祝贺/

(31b) /当即/冲/上/前/去/， /扭/住/一/名/歹徒/不/放/， /

(32) ‘上路’：作动词时不切分；‘上’作动词是要切分。

(32a) /我/背/起/你/的/薄被/送/你/上路/，

- (32b) /过去/村里/也是/上/路/打场/, /
- (33) ‘得了’: 取助词用法时不切分; 但作为‘动词+助词(了)’和可能式动补结构(见14)时要切分。
- (33a) /没/叫/到/你/的/时候/, /安心/等/着/就/得了/。
- (33b) /经/医生/诊断/他/得/了/胃癌/。
- (33c) /书记/何以/承受/得/了/, /
- (34) ‘得出’: 作动词时不切; 作可能式动补结构(见14)时要切分。
- (34a) [L 天津]/近几年/的/实践/得出/了/肯定/的/答案/。
- (34b) 为了/让/[dat 今年]/蒜农/的/产品/卖/得/出/、/卖/出/好/价钱/, /
- (35) ‘人称’: 作名词时不切分; 当‘称’作动词是要切分。
- (35a) 作者/用/第一/人称/的/叙述/手法/, /
- (35b) 据/用/过/的/人/称/, /打/国际/长途/如/从/[L 北京]/到/[L 美国]/, /每/分钟/只需/传统/电话/费用/的/[$\frac{1}{4}$]/。
- (36) ‘同行’: 用作名词时不切分, 读作‘tonghang’; 用作动词时读作‘tongxing’, 一律切分。
- (36a) /这时/恰/有/同行/到来/, /只好/借/[mon 一元钱]/给/他/。
- (36b) /笔者/与/她/骑车/同/行/。
- (37) ‘从小’:
- (37a) /图文/并茂/、/声/形/兼备/的/写作/能力/将要/从小/培养/, /
- (37b) /企业/从无到有/, /从/小/到/大/, /
- (38) ‘中学’:
- (38a) /在/长期/的/中学/教学/实践/中/我/体会/到/, /
- (38b) /引导/他们/在/实践/中/学/会/正确/行使/民主/权利/。
- (39) ‘上门’:
- (39a) /营业员/们/便/主动/上门/收款/。
- (39b) /打/出/了/名气/, /找/上门/来/的/工程/一个/接/一个/。
- (39c) 我/冲/出/门/去/, /随手/拉/上门/。/
- (40) ‘声响’: 作名词使用时不切分。
- (40a) /而/轻轻/地/挪动/椅子/走开/, /无/一点/声响/。/
- (40b) /“哗哗”/的/潮水/声/响/成/一片/, /
- (41) ‘就此’: 作副词使用时不切分。
- (41a) /国际/足球/界/一些/有识之士/就此/产生/一种/忧虑/, /
- (41b) 我们/也/欢迎/科技界/人士/就/此/问题/发表/意见/和/建议/。
- (42) ‘高层次’作形容词时不切分; 当该词前有副词修饰时需切分。
- (42a) /着眼点/放/在/培养/造就/大批/高层次/科技/人才/上/。

- (42b) /实现/更/大/规模/、/更/高/层次/的/扩张/和/发展/。/
- (43) ‘有的’：
(43a) /有的/用/汉语/，/有的/用/俄语/，
(43b) /是/[L 北大荒]/独/有/的/风味/。
- (44) ‘的话’：作助词使用时不切。
(44a) /如果/要/使/谈判/取得/迅速/进展/的话/，
(44b) /[P 卡比拉]/先生/对/我/的/话/是/持/认真/态度/的/。
- (45) ‘话说’：整体作动词使用时不切。
(45a) /话说/当年/，/他/言语/铿锵/：“/在/当时/，/一切/都/得/打破/常规/。
(45b) 用/他/自己/的/话/说/，
- (46) ‘标本’：意思为‘生物/标本’时不切；表示“直接和根本”并列的意思时要切开。
(46a) /他们/还/结合/挂图/、/标本/进行/讲解/。
(46b) /反/腐败/要/坚持/标/本/兼/治/，
- (47) ‘上将’：作为军衔使用时不切分。
(47a) /[O 中央军委]/副/主席/、/国务委员/兼/[O 国防部]/长/[P 迟浩田]/上将/
(47b) /[L 中国]/在/人口/问题/上/将/面临/新/的/挑战/。
- (48) ‘将军’：作为军衔使用时不切分。
(48a) /党/和/国家/领导人/、/解放军/元帅/、/将军/、/政府/省/部级/干部
(48b) /将/军：将/军/体/与/群体/紧密/结合/，/开办/体育/知识/讲座
- (49) ‘之一’：
(49a) /企业/领导班子/不/适应/社会主义/市场经济/的/要求/是/主要/原因/之一/。
(49b) /游人/视线/随/之/一/收/，/“/[L 太和宫]”/[int 三个]/大字/豁然/在/目/。/
(49c) /我/不禁/为/之/一/震/。
- (50) ‘到家’：作为形容词不切分。
(50a) /现在/不行/，/你/技术/不/过关/，/说明/练/得/还/不/到家/，/
(50b) /果/不/其/然/，/此/儿/到/家/就/猝/不/及/防/地/给/了/他/妈/一/刀/。
(50c) /[P 赵匡胤]/终于/将/义/妹/[P 京娘]/送/到/家/。
- (51) ‘在家’：
(51a) /一直/在家/等待/厂子/通知/上班/的/她/再/也/沉/不/住/气/了/
(51b) /实现/访客/在/家/门口/与/住户/可/视/
(51c) /把/她/一/人/放/在/家/中/[P 孙威锋]/放心/不/下/，
- (52) ‘人均’：
(52a) /学生/拥有/计算机/的/人均/占有/率/最高/
(52b) /[int 两]/人/均/未/达到/[fra 2 / 3]/的/当/选/票/数/，

(53) ‘中用’:

(53a) /“察/古/知/今/”/基本上/不/中用/了/，

(53b) /天文学/上/把/[L 宇宙]/中/用/光学/方法/看/不/到/的/物质/称/做/暗/物质/，

(53c) /西/体/[L 中]/用/我/也/反对/，

(54) ‘前去’:

(54a) /让/[P 胡洁青]/前去/扶持/、/帮忙/。

(54b) 一/名/应邀/到会/的/[L 北京]/小学生/激动/地/跑/上/前/去/请/他/签名/。

(55) 词表词“受过”只有“代人受过”的意思。当动词‘受’和助词‘过’构成“动+助”结构时，一律切开。

(55a) /它们/代/四/奸/受过/，

(55b) /[P 鲁迅]/虽然/在/[dat 二十年代中期]/受/过/[P 托洛茨基]/的/一定/影响/，

(56) ‘结果’：有名词和动词两种用法，都不切分，动词‘结果’的意思是杀死，而不是结出果实的意思。作为后一个意思，名词‘果’是动词‘结’的宾语，所以需切分。

(56a) /矫枉/过正/的/结果/，/是/大家/几乎/忘/了/怎么/吃/，/

(56b) /种/果树/一般/要/三年/才/能/结/果/，

9.2.4 ‘就是、只有、只是、还是’的切分规则

9.2.4.1 就是

‘就是’作‘副词、连词、助词’使用时不切分。但作‘动词’时，‘就’是副词，‘是’是动词，一律切分。

(A) 作副词时，共有 6 个义项：

(i) 单用，表示同意，对；

(ii) 表示坚决，不可更改；

(iii) 强调肯定某种性质和状态，含有反驳意味；

(iv) 强调迅速果断；

(v) 确定范围，排除其它；

(vi) 表示没有别的情况；。

(1) 我/一定/办到/，您/放心/就是/。/

(2) /反正/姥爷/就是/看/我/不/顺心/，/一点/也/不/喜欢/我/。/

(3) /望/着/车/来/车/往/的/马路/，/一/站/就是/[int 几个小时]/。/

(4) /就是/节目/诉求/为/非常/鲜明/的/单一/主题/，

(B) 作连词有 2 个义项：

(i) 表示假设的让步；即使(后面常用‘也’作呼应)；

(ii) 表示一种极端情况；纵然。如：

(5) 不是/播种/，/就是/锄地/；/不是/下/田/挖/野菜/，/就是/上山/打柴/。

(6) 这个/建议/好/倒/是/好/，/就是/远水/不解/近/渴/。/

(C) 就是作动词时，一律切分，‘就’是副词，‘是’是动词。如：

- (7) / [O 光华国中] / 职员 / [P 杨一中] / 就 / 是 / 买 / 菜 / 变成 / [O 慈德] / 会员 / 的 / 一 / 例 / 。
- (8) / 多元化 / 的 / 意思 / 就 / 是 / 有 / 了 / 更 / 多 / 的 / 选择 / 。
- (9) / 最 / 特别 / 的 / 就 / 是 / 黄金 / 压制 / 的 / 邮票 / 。
- (10) / 最 / 关键 / 的 / 原则 / 就 / 是 / 「 / 避 / 凶 / 趋 / 吉 / 」 / 。
- (11) 这 / 就 / 是 / [L 海尔-波普彗星] / 。

9.2.4.2 ‘只有’

‘只有’作为一个词表词有副词和连词两个义项。当他用作动词时，一律切开。

(A) ‘只有’做副词相当于‘只好’，表示唯一的选择。如：

- (1) / 家属 / 最后 / 只有 / 寄 / 望 / 对 / 岸 / [O 海协会] / 能 / 请 / [L 大陆] / 渔船 / 协 / 寻 / 。
- (2) / [L 中国] / 的 / 体育 / 长期 / 是 / 国家 / 一 / 家 / 办 / 的 / 。
- (3) 无 / 雪 / 的 / [dat 冬天] / 是 / 难 / 挨 / 的 / 。
- (4) 协办员 / 和 / 见习员 / 在 / 通过 / [int 三道] / 关 / 后 / 。
- (5) 如果 / “ / 邪恶 / 的 / 敌人 / 对 / [L 伊] / 发动 / 侵略 / 。

(B) “只有”作连词用表示必要条件，下文常用副词‘才、方’呼应。如：

- (6) 只有 / 掌握 / 了 / 最 / 先进 / 的 / 科学 / 。
- (7) 高尚 / 的 / 世界 / 只 / 对 / 高尚 / 的 / 人们 / 存在 / 。

(C) “只有”用作动词时一律切开。这时“只”做副词、“有”作动词。如：

- (8) / 完成 / 管理 / 的 / 比率 / 只 / 有 / [per 百分之八十九] / 。
- (9) / 车行 / 时速 / 只 / 有 / [len 卅到五十公里] / 左右 / 。
- (10) / 目前 / [O 基隆邮局] / 只 / 有 / 一个 / 集邮 / 柜台 / 。
- (11) / 因 / [O 中崙国小] / 整个 / 学区 / 只 / 有 / 一个 / [L 中崙里] / 。

9.2.4.3 ‘还是’

‘还是’有三种用法：连词、副词和动词。作动词时一律切分。

(A) 作连词用时表示选择，通常跟‘无论、不管’等连用。带连词‘还是’的句子，除疑问句外，‘还是’都可以换成‘或者’，意思不变。例如：

- (1) 无论是 / 说 / 新 / 话 / 。
- (2) 农民 / [P 张戎梅] / 说 / 。
- (3) 不管 / 是 / 开工 / 还是 / 竣工 / 。
- (4) 他们 / 不但 / 是 / 我们 / 公司 / 发展 / 的 / “ / 动力 / 之 / 源 / ” / 。

(B) ‘还是’作副词用时有三个义项：

- (i) 表示行为、动作或状态继续保持不变，相当于“仍然”、“依然”。

(ii) 表示经过比较后做出的选择。

(iii) 加强语气，相当于“到底”、“究竟”、“毕竟”。

‘还是/d’用在动词、形容词前，可以省作‘还’，而用在主语前不能省作‘还’。

- (5) /但/现场/交通/还是/十分/杂乱/。
- (6) /该/基金/还是/可以/支应/灾民/最高/[mon 一百万元]/的/贷款/额/，
- (7) /很多/居民/还是/使用/地下水/，
- (8) /[P 陈]/还是/不/改/顽皮/个性/，
- (9) /多数人/还是/喜欢/为/宝宝/选/个/金/饰/，
- (10) /[P 陈小弟]/的/父/母亲/还是/勇敢/地/生/下/他/，

(C) ‘还是’用作动词时一律都要切开，即‘还’作副词使用，‘有’作动词使用。句型“是/v.....的”可以帮助我们判断‘还是’在句中是不是一种动词的用法。

- (11) /关键/还/是/在/府/会/双方/态度/，
- (12) /初/到/部队/，/[age 十五六岁]/，/还/是/个/没/见/过/世面/的/毛孩子/。/
- (13) /她/不/相信/歌剧/这门/综合/艺术/会/落入/低谷/，/认为/关键/还/是/提高/歌剧/自身/的/品质/。/
- (14) 但/在/日常/工作/中/，/我/深感/除了/忙/还/是/忙/，/搞/得/焦头烂额/，/一天到晚/自己/不/属于/自己/。/

9.2.4.4 ‘只是’

‘只是’有三种用法：副词、连词和动词。作动词时统统切分。

(A) ‘只是’作副词使用时有两个义项：

(i) 表示限定某种情况或范围，相当于‘仅仅是’。句末用‘而已’或‘罢了’等配合，表示语气更为缓和。

(ii) 强调在任何条件下情况都不变，有‘总是’的意思。

- (1) /只是/作为/预定/分娩/日/的/参考/。
- (2) /只是/没有/焯/窑/经验/的/[P 张]/课/长/，
- (3) /他/虽/表示/民意调查/结果/数据/只是/具有/参考/价值/，
- (4) /施工/初期/只是/修剪/树枝/，

(B) ‘只是’作连词用，用在后一分句，表示轻微的转折，补充修正上文的意思，与‘不过’的用法相近。

- (5) /记者/在/重灾区/[L 大河乡]/注意/到/，/群众/有/饭/吃/，/有/衣/穿/，/有/伤病/能/医治/，/只是/搭建/的/小/窝棚/难以/抵御/坝上/呼啸/的/寒风/。/
- (6) [dat 唐朝]/著名/诗人/[P 李商隐]/『/夕阳/无限/好/，/只是/近/黄昏/』/的/诗句/是/对/黄昏/的/叹息/和/无奈/，/

(C) ‘只是’用作动词时一律要切开，即‘只’作副词，‘是’作动词。如：

- (7) /他/只/是/[dur 一个月]/领/[mon 二万多元]/的/工人/
- (8) /事实上/[L 盐埔乡]/公所/的/薪水/无/着落/只/是/冰山/一/角/，
- (9) /这些/需求/不/只/是/钱/或/资源/，/