

中文电子病历词性标注规范

(草案)

《中文电子病历词性标注规范》是根据以下资料提出的：

1. 《信息处理用现代汉语词类标记规范》，中国国家标准 GB/T 20532-2006
2. 《973 当代汉语文本语料库分词、词性标注加工规范》(草案)
山西大学计算机科学系 山西大学计算机应用研究所
3. 《The Part-Of-Speech Tagging Guidelines for the Penn Chinese Treebank》(v 3.0)
University of Pennsylvania, 2000

一、词性标注总则

信息处理用电子病历词性标注主要原则有**三个**：

- (1) 语法功能原则。语法功能是词类划分的主要依据。词的意义不作为划分词类的主要依据，但有时也起着某些参考作用。
- (2) 允许有兼类。根据各种统计研究，电子病历的某些词具有多种语法功能，但这多种功能的分布概率不同。在信息处理用电子病历词类体系中，各词类的确立要根据词的主要语法功能。

二、词类标记集

本规范的词类标记集采用《The Part-Of-Speech Tagging Guidelines for the Penn Chinese Treebank》的大类，只针对电子病历的上层应用（例如信息抽取）增加了部分细类。

词类分别为：

名词	NT	时间名词
	NR	专有名词
	NN	普通名词
动词	VC	系动词
	VE	“有”字动词
	VV	普通动词
形容词	VA	表语形容词

30个类别

	JJ	名词修饰语
数词	CD	基数词
	OD	序数词
	DT	限定词
量词	M	/
副词	AD	/
代词	PN	/
介词	P	/
定位词	LC	/
连词	CC	并列连词
	CS	从属连词
助词	DEC	补语标记/名词化标记 (de5)
	DEG	属格标记/关联标记 (de5)
	DER	结果助词 (de5)
	DEV	方式助词 (de5)
	SP	句末助词
	AS	动态助词
	ETC	“等”字助词
	MSP	其他助词
其他词	LB	长“被”结构词 (bei4)
	SB	短“被”结构词 (bei4)
	BA	“把”结构词 (ba3)
	PU	标点符号

三、词性标注详细说明

1. 名词

表示人和事物的名称或时间、处所等，在句中主要充当主语和宾语。

名词通常是谓语或介词的论元：

- 1) 名词不能被程度副词或否定副词修饰
- 2) 许多名词能被[限定词+量词]修饰

3) 名词能直接修饰名词

如果某词是名词短语的中心词，则标注为名词。

名词短语的判定依据：

1) 如果某短语能被[限定词+量词]修饰，并且[限定词+量词]在其他上下文中仅修饰名词短语，则判定该短语为名词短语。

2) 如果某短语是动词或介词的论元，并且在其他上下文中仅充当名词短语论元，则判定该短语为名词短语。

3) 如果某短语被“的”修饰，那么该短语很可能是名词短语

1.1 普通名词 (NN)

表示人和事物的名称。除专有名词(NR)、时间名词(NT)、定位器(LC)之外的所有名词。一般不能直接修饰动词，也不能通过添加“地”修饰动词。

1.1.1 基本式

常用的非合成名词，例如：

豆包#NN

衣服#NN

药物#NN

高血压#NN

肝#NN

1.1.2 合成式

1.1.1.1 并列关系

深浅#NN

言语#NN

血尿#NN

口眼#NN

1.1.1.2 定中关系

[1] [名+名]

脑实质#NN

腱反射#NN

肌张力#NN

高血压病#NN

精神状态#NN

神经内科#NN

[2] [动十名]

饮酒史#NN

呕吐物#NN

烧心感#NN

硬化症#NN

[3] [数十名]

半球#NN

四肢#NN

1.1.1.3 主谓关系

[名+动]

癌变#NN

脑出血#NN

脑梗塞#NN

1.2 时间名词（NT）

表示时间。可做介词的宾语，可被替代为“这个时候”，或者用“什么时候”提问，也可直接修饰动词短语或句子。

1.2.1 基本式

[1] 一周的七天，农历的初一到初十，“（大）年初一”到“（大）年初十”，例如：

星期一#NT 初三#NT 年初二#NT 大年初一#NT

[2] 年月日时分秒，按年、月、日、时、分、秒标注为 NT，例如：

1997 年#NT 3 月#NT 19 日#NT，

下午#NT 2 时#NT 18 分#NT 35 秒#NT

[3] “前、后、上、下、大前、大后、头”加“天”或“上 / 下”加“月 / 周 / 星期”时，标注为 NT，例如：

前天#NT

上周#NT

下星期#NT

[4] 数字与“:”或“-”结合在一起的表示具体时间的串，整体标注 NT，例如：

08:35:28#NT 2003-03-29#NT

1.2.2 合成式

[1] [代+定]

此后#NT

[2] [名+定]

目前#NT

今后#NT

1.3 专有名词（NR）

某个人，地理位置，机构等专有的名称。包括：区域/国家/县/市，山/河，报纸/杂志，组织/公司，学校/社团/基金会，个人/家庭成员。

以下词不是 NR：国籍(例：中国人)、种族(例：白人)、职位头衔(例：教授)、疾病、职业、器官(例：肺)、乐器(例：钢琴)、运动(例：足球)、花(例：玫瑰)

1.3.1 人名

表示人的名称（姓名）

汉族或类汉族人名，整体标注 NR，例如：

田焕章#NR 刘淑芬#NR

1.3.2 地名

表示地理区域的名称

[1] 地名后有“省、市、县、区、乡、镇、村、旗、州、都、府、道”等单字的行政区划名称时，标注为 NR，例如：

黑龙江省#NR 哈尔滨市#NR 青冈县#NR

[2] 地名后有表示自然区划的一个字的普通名词，“街，路，道，巷，里，町，庄，村，弄，

堡”等，标注为 NR，例如：

太阳村#NR 兴安街#NR

1.3.3 族名

表示民族或部落名称的专有名词，例如：

汉族#NR 回族#NR 朝鲜族#NR

1.3.4 团体机构名

包括团体、机构、组织的专有名称

[1] 团体、机构、组织的专有名称若作为名词登录在分词库中，则直接标注为 NR，其缩略形式标注为 NR，例如：

哈医大#NR

[2] 大多数团体、机构、组织的专有名称一般是短语型的，较长，且含有地名或人名等专名，对于词表中没有收录的，暂不做整体标注，例如：

哈尔滨#NR 医科#NN 大学#NN

2. 动词

表示动作、行为，人或动物的心理活动、生理状态，事物的存现、变化等，在句子中主要充当谓语。

动词的判定依据：

- 1) 动词（不包括助动词）通常作为子句的谓语使用
- 2) 动词能够被“不”，“没”否定
- 3) 方向标记能够附属于大多数动词
- 4) 多数动词能够出现在[VV+ “不” +VV]结构中

注意：如果动词列表中的词是名词短语的中心词，则标注为名词；如果动词列表中的词修饰名词（不包括关系子句的中心词），则标注为名词或名词修饰语。

2.1 普通动词（VV）

表示动作、行为，人或动物的心理活动、生理状态等，

2.1.1 基本式

伸#VV

咯#VV

就诊#VV

2.1.2 合成式

2.1.1.1 并列关系

抬举#VV

2.1.1.2 动宾关系。

持物#VV

进食#VV

查体#VV

构音#VV

用药#VV

2.1.1.3 状中关系

[1] [形十动]

好转#VV

粗测#VV

[3] [动十动]

飞奔#VV

2.1.1.4 动补关系

[1] [动十结（动、形、趋）]

达到#VV

升高#VV

[2] [动十“于”]

低于#VV

2.1.3 能愿动词

表示可能、意愿。如：

能够#VV

肯#VV

可以#VV

可能#VV

情愿#VV

要#VV

会#VV

应该#VV

2.1.4 趋向动词

表示动词实际的和抽象的趋向。例如：

入#VV

2.2 “有”字动词（VE）

“有”作为表示所有格、存在的主要动词，例如：

有#VE

没有#VE

无#VE

2.3 系动词（VC）

表示关系的判断，如：

是#VC

为#VC

非#VC

注：“非”作“不是”的意思，并且句子中没有其他动词

3. 形容词

表示性质、状态，在句中主要充当谓语、定语、状语和补语。

3.1 表语形容词（VA）

表示性质。不接宾语，或后接“的”再修饰名词

[1]不接宾语，并且能被“很”修饰的谓词，例如：

蛋白:1 +#VA

水平眼震（-#VA）

运动 自如#VA

嗅觉 正常#VA

言语 笨拙#VA

[2]由[1]中的词通过叠字形成，例如：

红彤彤#VA

[3][名+形]，可表示成像…一样…，例如：

雪白#VA

[4]当形容词和名词之间有“的”，该形容词能被程度副词修饰，则标注为 VA，例如：

明显#VA 的眩晕

注意：如果属于 VA 中的词后接宾语，则标为 VV

3.2 名词修饰语 (JJ)

表示状态。直接修饰名词，或后接“的”再修饰名词

[1] 区别词，使用模式一般为 [JJ+ “的” +名] 或 [JJ+名]，该词不能被程度副词修饰，例如：

持续#JJ 的#DEG 时间#NN

右#JJ 手#NN

左下#JJ 肢#NN

[2]构造词，通常是长词的缩写形式，构成方式一般包括：[动+名]，[介+名#方位]，[副+表]，使用时直接修饰名词，例如：

留美#JJ 学者#NN

[3]直接修饰名词，不管形容词是否在 VA 中，只要形容词后没有跟“的”，该词就都标注为 JJ，例如：

习惯性#JJ 动作#NN

周围性#JJ 面瘫#NN

4. 数词

4.1 基数词 (CD)

[1] 基数（可接概数词，如“来”、“多”、“好几”等）、小数、分数、百分数标注为 CD，例如：

75#CD

3.0#CD

百#CD

[2] “好些”、“若干”、“半”、“多”等词，标注时注意：

1) 当修饰名词或量词时，标注为 CD，例如：

多#CD 次#M

2) [副+形]

充当谓语时，标注为 VA，例如：

汗液#NN 较#AD 多#VA

3) 在量词后充当程度副词，例如：

两#CD 年#M 多#AD

4.2 序数词 (OD)

表明顺序。[第+基数]或省略“第”是常见的形式，例如：

第二#OD

二#OD 院#NN

4.3 限定词 (DT)

在名词词组中对名词中心词起限定性作用的词类，限定中心词的作用范围，包括：

[1] 指示限定词，例如：

这#DT 次#M

[2] 其他数词（不包括基数词和序数词），例如：

各#DT 向#NN 运动#VV 自如#AD

每#DT 日#M 20#CD 支#M

任何#DT 药物#NN

整#DT 句#M 话#NN

所有#DT 后果#NN

一切#DT 后果#NN

5. 量词 (M)

表示事物的单位或动作的量。包括，常和名词连用的名量词，有个体量词（位、辆、张），度量词（克、千米），复合量词（人次、架次、吨公里），不定量词（点、些）；以及动量词（次、回、趟）和时量词（天、小时）等。

[1] 基本量词例如：

3#CD 个#M 月#NN

36.2#CD ℃#M

[2] 复合量词，例如：

次/分#M

mmol/l #M

7. 副词 (AD)

说明动作行为或状态性质等所涉及的范围、时间、程度、频率以及肯定或否定的情况，在句中
能且只能做状语。副词主要修饰动词和形容词。

[1] 连接副词，例如：

否则#AD

所以#AD

因此#AD

则#AD

随后#AD

然后#AD

例如#AD

[2] 时间副词，例如：

届时#AD

即将#AD

紧接着#AD

[3] 频率副词（直接修饰动词），例如：

多次#AD

[4] 程度副词，例如：

极为#AD

较#AD

较为#AD

[5] 方式副词，例如：

互利#AD

[6] 情态副词，例如：

将#AD

[7] 修饰数量的副词，例如：

近#AD

不足#AD

[8] 修饰名词短语的副词，例如：

又#AD

才#AD

[9] 否定副词，例如：

未#AD

没#AD

不#AD

[10] 短语词，例如：

进一步#AD

越来越#AD

尤其是#AD

据称#AD

据悉#AD

[11] 其他副词，例如：

一起#AD

另#AD

正在#AD

凡#AD

才#AD

8. 代词 (PN)

用来代替名词短语，表示先前指定的或上下文中的人或事。通常不能用限定词或形容词性表达式更替。包括人称代词、物主代词、指示代词、自身代词等。

[1] 人称代词，例如：

你#PN

它#PN

我#PN

之#PN

我们#PN

他们#PN

[2] 指示代词，例如：

这#PN

这儿#PN

那#PN

那里#PN

此#PN

这里#PN

[3] 物主代词，例如：

其#PN

[4] 反身代词，例如：

他自己#PN

自己#PN

[5] 其他代词，例如：

彼此#PN

大家#PN

对方#PN

双方#PN

自身#PN

9. 介词 (P)

引介名词性成分，不能单独充任句子成分。

可接名词短语或一个子句。

注意：“把”和“被”不标记为 P

单音节介词后加“着”仍是介词，标注为 P，例如：

为着#P

朝着#P

向着#P

对着#P

沿着#P

随着#P

其他介词包括：

[1] 动词式，例如：

经过#P

截止#P

有关#P

离#P

[2] 从属连词式，例如：

随着#P

沿着#P

鉴于#P

除了#P

为了#P

[3] 副词式，例如：

就#P

10. 定位词（LC）

界定时间的范围或地点的方位，定位词的功能就是前接名词短语或子句，整体作为介词的论元，或直接修饰动词短语或子句，LC 包括：

[1] 方位词，表明方向，位置等，不能被“限定词#数词+量词”结构修饰。

a) 单音节词，例如：

前#LC

后#LC

里#LC

外#LC

内#LC

北#LC

东#LC

边#LC

侧#LC

底#LC

间#LC

末#LC

旁#LC

b) 双音节词，例如：

之间#LC

以北#LC

c) 两个单音节词连用，例如：

前后#LC

左右#LC

上下#LC

东北#LC

[2] 其他词语，例如：

自#P 发病#VV 以来#LC

近#AD 1#CD 月#M 来#LC

到#P 目前#NT 为止#LC

从#P 四月#NT 开始#LC

11. 连词

连接词、短语或句子，表示两者之间所具有的某种逻辑关系。连词单独切分。

11.1 并列连词（CC）

连接两部分，每部分在整句中有相似功能。主要包括：

[1] “与”关系，例如：

与#CC

和#CC

及#CC

并且#CC

细胞浸润，而#CC 没有…

脑血流图 +#CC 颅外段检查

[2] “或”关系，例如：

或#CC

或者#CC

[3] 共现的并列连词，例如：

既#CC……又#CC

又#CC……又#CC

不仅#CC……而且#CC

[4] 其他并列连词：

数秒 至#CC 数分钟

可至 170 -#CC 180 mmHg

11.2 从属连词（CS）

连接两个子句，其中一个从属于另一个。例如：

因为#CS

虽然#CS

只要#CS

如果#CS

即使#CS

以便#CS

注意：连接副词经常出现在主句中，与从属连词匹配，例如：

如果#CS...就#AD...

12. 助词

11.1 补语标记/名词化标记（DEC）

仅包括“的”，“之”。

[1] 主语是一个子句，例如：

既往#NT 有#VE 栓子#NN 来源#NN 的#DEC 基础#NN 疾病#NN

[2] 主语是动词短语，例如：

可能#VV 出现#VV 的#DEC 副作用#NN

11.2 属格标记/关联标记（DEG）

仅包括“的”，“之”。

主语可以是名词短语，介词短语，形容词，限定词，例如：

其他#DT 脏器#NN 的#DEG 栓塞#NN

11.3 结果助词（DER）

仅包括“得”。

[动+“得”]，例如

吃#VV 得#DER 多#VA

11.4 方式助词（DEV）

仅包括“地”，“的”。

一般修饰动词短语，例如

明显#AD 的#DEV 加重#VV

高兴#VA 地#DEV 说#VV

11.5 句末助词（SP）

包括：了，呢，吧，啊，呀。

通常位于句末，例如：

他#PN 还#AD 好#VA 吧#SP ? #PU

可用来表暂停，例如：

他#PN 吧#SP ， #PU 人#NN 很#AD 好#VA 。 #PU

11.6 动态助词（AS）

表状态，仅包括“了”，“着”，“过”，“的”。

例如：

进行#VV 了#AS 14#CD 天#M

出现#VV 了#AS 右#JJ 手#NN 麻木#VA

11.7 “等”字助词（ETC）

包括“等”，“等等”。

常后接名词短语，或用于短语末尾，例如：

伴有#VV 风心病#NN 等#ETC 心脏病史#NN

11.8 其他助词（MSP）

包括“所”，“以”，“来”，“而”。

常后接动词短语，例如：

所#MSP 致#VV

以致#MSP

12. 其他词

12.1 长“被”结构词（LB）

仅包括“被”，“叫”，“给”，“为”。在长“被”结构中，同时存在施动者和受动者。

长“被”结构：[名+LB+名+动]，例如：

患者#NN 被#LB 家属#NN 发现#VV

12.2 短“被”结构词（SB）

仅包括“被”，“给”。在短“被”结构中，不存在施动者，仅有受动者。

短“被”结构：[名+SB+动]，例如：

患者#NN 被#SB 叫醒#VV

12.3 “把”结构词（BA）

仅包括“把”，“将”。

“把”结构：[名+BA+名+动]，例如：

患者#NN 把#BA 呛咳物#NN 咳出#VV

12.5 标点符号（PU）

注意：当标点是其他词的组成部分时，不进行标注，例如：

10,000#CD 元#M

注意：

为区别标签和文本，待标注的括号需要进行以下替换：

- (1) “(” 和 “)” 替换成 “-LRB-” 和 “-RRB-”
- (2) “{” 和 “}” 替换成 “-LCB-” 和 “-RCB-”
- (3) “[” 和 “]” 替换成 “-LSB-” 和 “-RSB-”

三、电子病历标注歧义说明

1. 动名词标注说明

如果词是名词短语的中心词，标注为 NN；如果是动词短语的中心词，标记为 VV。

标注方法：

- 1) 如果该词被[限定词/数词+量词]修饰，标注为 NN。
- 2) 如果该词是某从句的谓语，并且该从句修饰名词，则该词通常标注为 VA 或 VV，否则标注为 NN。
- 3) 如果该词被[短语+“地”]修饰或者被副词修饰，则标注为 VV。
- 3) 当该词是某动词论元的中心词时，如果论元为名词短语，则该词标注为 NN，例如：

进行对宗教的 改革#NN

如果论元为动词短语或子句，则该词标注为 VV，例如：

进行 改革#VV 宗教

补充说明：上述方法无法消歧的情况下，**选择词语的本质属性标注**。

例如“伴有视物模糊”中的“视物”。从描述症状角度看，“视物”可以标注为 NN，从短语结构看，“视物模糊”又是动补短语，我们将这种情况统一标注为动词。

2. 特殊符号标注说明

中文电子病历中，特殊符号作为一种缩略词被频繁使用，然而 PCTB 规范中并没有相关说明，“+”甚至没有在 PCTB 语料库中出现过，只能被标注为标点符号(PU)。我们规定应该**根据上下文标注特殊符号**，例如“肌力 4+级”中“+”表示“强”，应标为形容词 (VA)，“头 MRI+MRA 示”中“+”则表示“和”，标为并列连词(CC)。

3. 省略成分标注说明

名词修饰语(JJ)的标注歧义在查体部分经常出现，例如“左侧肢体麻木”和“左侧中枢性面瘫”中的“左侧”。在 PCTB 规范中“左侧”可能词性包括名词(NN)、定位词(LC)以及名词修饰语(JJ)，其中 LC 可以根据是否为介词论元或直接修饰动词短语(子句)来判定，NN 和 JJ 则需要判断该词是否成为中心词，可是在上例中“左侧”都不是中心词，无法达到消歧的目的。第一个例子是正常语法结构，“左侧”做 JJ 直接修饰“肢体”，而第二例子实际上省略了谓语“存在”，补充完整后“左侧”应为名词主语，所以针对类似问题，我们规定**首先补充语法成分再进行标注**。

四、其他标注歧义说明

1. 副词(AD) or 动态助词(AS)

“在”位于动词前时是 AD 不是 AS。

注意：标为副词时，“在”和动词之间可以插入其他副词；做 AS 时不能在动词和“在”之间插入其他词。

2. 副词(AD) or 并列连词(CC)

AD 和 CC 的兼类词有：“又”，“还是”。CC 可以连接对等的两个成分，AD 不可以。例如：你咋 又#AD 来了。

这本书既实用 又#CC 有趣。

3. 副词 (AD) or 名词修饰语 (JJ)

后接名词短语时，可以是 AD 或 JJ。

没有 AD 和 JJ 的兼类词。一些 AD 也可以修饰名词，比如：“才”，“又”。区分 JJ 和 AD 的方法是在该词和名词短语之间插入“的#DEG”，如果依然合法，为 JJ，否则为 AD。

4. 副词 (AD) or 普通名词 (NN)

后接动词短语时，是副词或是名词。

时间名词和方位名词可直接修饰动词短语或句子，这种情况下仍为名词。

除上述两种名词外，如果在该词之前添“在”不改变其原意或句子的正确性，该词为名词，否则为副词。例如：

政治#AD 解决这个问题。

重点#AD 抓生产。

5. 副词 (AD) or 时间名词 (NT)

后接动词短语时，是副词或是时间名词。

如果该词能作为名词短语的开头，则为 NT，否则为 AD。

例如：

早日#AD 实现

目前#NT

今后#NT

6. 副词 (AD) or 表语形容词 (VA)

1) 规定 VA 不能直接修饰动词短语，需要接上“地#DEV”。

2) 后接“地”再修饰动词短语：如果标为 AD 或 VA 有相同的含义，则标记为 VA，否则，根据上下文确定。

3) 直接修饰动词短语：标记为 AD。

例如：

大大#AD 提高

高兴#VA 地……

紧密#VA 地……

7. 副词 (AD) or 普通动词 (VV)

后接动词短语或句子时，可以是 AD 或 VV

动词可以使用[VV+ “不” +VV]的结构。

容易混淆的词：

情态式副词：大概#AD，将#AD，一定#AD

副词式动词：是否#VV

8. 动态助词 (AS) or 普通动词 (VV)

前接动词时，可以是 AS 或 VV

注意：AS 仅有四个：“着”、“了”、“过”、“的”

容易混淆的词：

结构[VV+R]中，一些 R 与 AS 有类似功能，此时标注为 VV。

例如：

跑 起来#VV

写 完#VV

注意：这两个例子中的“起来”和“完”，容易被错误标注为 AS。

9. 并列连词 (CC) or 介词 (P)

“与”、“和”、“跟”、“同”，这些词既是 CC 也是 P。

在该词处于两个名词短语中间时，可以是 CC 或 P。

标注方法如下：

1)如果两名词短语可以交换，则是 CC。

2)如果该词前可以被副词修饰，则是 P。

3)如果句子更强调前一个名词短语，则是 P。

4)如果去掉前一个名词短语，后面也必须被去掉，则是 CC。

例如：

我和#CC 她是好朋友。

和#P 她一样，我也很喜欢英语。

10. 从属连词 (CS) or 介词 (P)

CS 后只能接句子，而 P 后可以接句子或名词短语。

CS 和 P 都可以接子句，[CS +子句 1+ “，” +子句 2]和[P +子句 1+ “，” +子句 2]区别如下：

1) [P +子句 1+ “，” +子句 2]中，[P+子句 1]可以插入到子句 2 的主语和谓语之间。

[CS +子句 1+ “，” +子句 2]则不能。

例如：

对#P 我不能按时来，他感到很失望。

替换后：

他对#P 我不能按时来，感到很失望。（合法）

虽然#CS 不能去，你还是应该去。

替换后：

你虽然#CS 不能去，还是应该去。（不合法）

2) [CS +子句 1+ “，” +子句 2]中，CS 可以插入到子句 1 的主语和谓语之间。[P +子句 1+ “，” +子句 2]则不能。

例如：

虽然#CS 他不能上场，这个队还是会赢。

替换后：

他虽然#CS 不能上场，这个队还是会赢。（合法）

对#P 我不能按时来，他感到很失望。

替换后：

我 对#P 不能按时来，他感到很失望。（不合法）

3) [P +子句 1+ “，” +子句 2]中，子句 1 可以被名词短语代替。[CS +子句 1+ “，” +子句 2]则不能。

例如：

对#P 我们不能按时来，他感到很失望。

替换后：

对#P 这件事，他感到很失望。（合法）

虽然#CS 我不能去，你还是应该去。

替换后：

虽然#CS 这件事，你还是应该去。(不合法)

11. 基数词 (CD) or 限定词 (DT)

CD 中的词是可以定量的。DT 除了含有一些不可定量的数词外，还含有一些指示限定词。

容易混淆的词：

CD：许多，若干，个把。

DT：各，全，某，这，全体，全部，一切

指示限定词：这，这些，此，该，本，那，那些，上，下，前，后，头，另，其余，其他，其它，某，某些

12. 基数词 (CD) or 名词修饰语 (JJ)

CD 和 JJ 都可以修饰名词。JJ 描述的是其修饰名词的属性，该属性可以区分出全体中的某个个体。CD 是一些数词，代表所修饰名词的数量。

CD：一些，大批，好几，不少。

JJ：共同，女。

13. 基数词 (CD) or 时间名词 (NT)

当数词有时间的含义时，标注为 NT。

例如：

1991#NT 至 1995 年。

14. 限定词 (DT) or 名词修饰词 (JJ)

JJ 描述了所修饰名词的属性，而 DT 说明所修饰名词的限定范围。

例：

美国 前#JJ 总统。

前#DT 7 个月。

15. 限定词 (DT) or 序数词 (OD)

“首”是 OD，不是 DT。

16. 限定词 (DT) or 代词 (PN)

指示限定词(例：“这”、“那”)是 DT 和 PN 的兼类词。

标注方法：

1) 词语单独做主语时，标注为 PN；后接“的”再修饰名词短语时标注为 PN。

2) 词语直接修饰名词短语时，标注为 DT。

3) DT 中除指示限定词之外的所有词，不管以上述哪种结构出现，都标注为 DT。

4) PN 中除了指示限定词之外的所有词，不管以上述哪种结构出现，都标注为 PN。

指示限定词：这，这些，此，该，本，那，那些，上，下，前，后，头，另，其余，其他，其它，某，某些

17. 名词修饰语 (JJ) or 普通名词 (NN)

JJ 不能作为名词短语的中心词，NN 则可以。

规定按照下列方式构成的词，视为 NN：NN+形，NN +状，NN +制。

NN +形：椭圆形，V 形

NN +状：带状，颗粒状

NN +制（当‘制’表示‘制度’的时候）：货币制，股份制

18. 名词修饰语 (JJ) or 介词 (P)

“有关”是 JJ 和 P 的兼类词。

例如：

有关#JJ 单位

有关#P 撤军的报告

19. 定位词 (LC) or 其他助词 (MSP)

“来”是 LC 和 MSP 的兼类词。

例如：

近年 来#LC (界定范围)

用 暴力 来#MSP 维持 (表目的)

20. 定位词 (LC) or 普通名词 (NN)

LC 通常附属于名词短语或子句，所以整个部分都是介词的论元，或者直接修饰动词短语或子句。

20. 量词 (M) or 普通名词 (NN)

前接基数词 (CD)：如果可以在 CD 和该词之间插入量词，并不改变句子的含义，则标记为 NN。否则，标记为 M。临时量词指的是，由名词充当的临时的量词，可以修饰名词。

例如：

一#CD 学生

一#CD 年

一#CD 箱子书

21. 普通名词 (NN) or 专有名词 (NR)

专有名词是地名，人名，城市名等等。一般不能被[限定词+量词]修饰。

以下词结尾的词标注为 NN：方，军，人，会，队，厅。

例如：美方#NN，美军#NN，美国人#NN，办公厅#NN

22. 普通名词 (NN) or 时间名词 (NT)

一个 NT 为首的名词短语可以直接修饰句子或动词短语并可回答问题“什么时候”。除了“地点”名词，其他名词不能直接修饰句子或动词短语

23. 普通名词 (NN) or 表语形容词 (VA)

如果该词可以被“很”修饰，则把它标注为 VA。

例如：

表示 乐观#VA

24. 普通名词 (NN) or 普通动词 (VV)

如果词是名词短语的中心词，标注为 NN；如果是动词短语的中心词，标记为 VV。

标注方法：

1) 如果该词被[限定词/数词+量词]修饰，标注为 NN。

2) 如果该词是某从句的谓语, 并且该从句修饰名词, 则该词通常标注为 VA 或 VV, 否则标注为 NN。

3) 如果该词被[短语+“地”]修饰或者被副词修饰, 则标注为 VV。

3) 当该词是某动词论元的中心词时, 如果论元为名词短语, 则该词标注为 NN, 例如:

进行对宗教的 改革#NN

如果论元为动词短语或子句, 则该词标注为 VV, 例如:

进行 改革#VV 宗教

25. 普通名词 (NN) or 代词 (PN)

[限定词+名词]或[代词+名词]既有代词的特性, 又有名词短语的特性。出于一致性, 我们把它们全部视为 NN 或 NT。

NN: 本人#NN, 本校#NN, 全球#NN, 当地#NN, 我校#NN

NT: 当今#NT, 何时#NT

26. 介词 (P) or 普通动词 (VV)

标注方法:

1) 如果整个句子中没有其他的动词, 标为 VV。

例如:

他 在#VV 家。

2) 如果该结构中缺少宾语 (短回答或是[动词短语+“不”+VV]问句), 标注为 VV。

例如:

他 在#VV 家 不#AD 在#VV ?

3) 如果词后可以跟动态助词, 则为 VV。注意: 一些介词以动态助词结尾,

例如:

为了#P 。

4) 一些词是 P 和 VV 的兼类词, 但意思有些不同:

例如:

靠#P(意思为: 倚着), 靠#VV(意思为: 依靠、依赖):

-他 靠#P 墙 站着。

-你 不 能 总 靠#VV 父母 。

5) 如果经过上述方法, 词性依然不清楚, 则标注为 P。

27. 表语形容词（VA） or 普通动词（VV）

VA 不能接宾语，所有的 VA 除了原本包含了“绝对”“很”含义的 VA 外，其他都可以被“很”修饰。

一些词是 VA 和 VV 的兼类词，比如：丰富，少，多。

例如：

这儿 少#VV 了#AS 一本书，这本书 很少#VA。

注意：宾语需要进行界定：

1) 给定一个动词和一个名词短语，如果[介词+名词短语+动词]和[动词+名词短语]都是符合语法规则的，则将该名词短语看做动词的宾语，并将该动词标注为 VV。

例如：

他 对 我 很 关心#VV。

2) 如果动词后边跟着名词短语或是被[限定词+量词]修饰，来表明程度，那么，名词短语或[限定词+量词]不是动词的宾语。

例如：

鞋 大#VA 了#AS 一号。

28. “有”字动词（VE） or 普通动词（VV）

“有”、“没”、“没有”、“无”，这些词做主要动词时，标记为 VE。

容易混淆的词：“具”、“具有”、“拥有”、“富有”，这些词虽然含有“有”，但仍标注为 VV。

29. 特殊词语

1) 的

有四种标记：DEC、DEG、AS、SP。

例如：

DEC: 我 买 的#DEC 书。

DEG: 我 的#DEG 书。

SP: 他 是 一 定 要 来 的#SP。

AS: 他 是#VC 在 这 里 下 的#AS 车。

2) 等、等等

修饰前接短语时，被标价为 ETC

例如：

科技、文教 等#ETC 领域

科技、文教 等等#ETC

“等”还有其他的标记：M、VV

例如：

二 等#M 奖学金

他 会 等#VV 你 的#SP 。

3) 来

有 VV、SP、MSP、LC 等标记，也可以是 CD 的一个组成部分

例如：

VV:

他 来#VV 了#AS 。

他 拿来#VV 了#AS 一 本 书。

他 走#VV 上来#VV 了。

SP:

他 拿出 书 来#SP 。

MSP:

他 用 这 个 来#MSP 证明 他 是 无 辜 的。

LC:

3 年 来#LC

CD 的一部分:

30 来#CD 个#M

4) 连

在“连……都#也”结构中，标记为 AD。

例如：

他 连#AD 我 都 不 认识。

5) 有

“有”充当主要动词是被标注为 VE。

“没有”表示“不拥有”的时候，单独做谓语，标记为 VE；当表示“没”，修饰动词时，标记为 VV。

例如：

他 没有#VE 这 本 书。

他 没有#VV 来。

“没”独立做主要动词时，标注为 VE；当修饰动词短语或出现重复词中间时，标记为 AD。

例如：

他 没#VE 书。

他 没#AD 来#VV。

他 来#VV 没#AD 来#VV？

“有没有”，做主要动词时标记为 VE（有#VE 没有#VE）；当修饰动词短语时，标记为 VV（有#VV 没有#VV）。

6) 这样

有 PN、AD、VV 等标记。

例如：

这样#PN 的#DEG 伙伴#NN 关系#NN

你 这样#AD 做#VV 不 对。

你 别 这样#VV。