**词性标注文档**

1. **汉语中的词类及其标注表示**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 代码 | 名称 | 子类  标记 | 特殊用  法标记 |  | 代码来源 |
| a | 形容词 |  |  |  | 取英语形容词adjective首字母 |
| b | 区别词 |  |  |  | 取汉字“别”的声母 |
| c | 连词 |  |  |  | 取英语连词conjunction首字母 |
| d | 副词 |  |  |  | 取英语副词adverb第二个字母 |
| e | 叹词 |  |  |  | 取英语叹词exclamation首字母 |
| f | 方位词 |  |  |  | 取汉字“方”的声母 |
| g | 语素 |  |  |  | 取汉字“根”的声母 |
| h | 前接成分 |  |  |  | 取英语head首字母 |
| i | 成语 |  |  |  | 取英语俗语idiom首字母 |
| j | 简缩略语 |  |  |  | 取汉字“简”的声母 |
| k | 后接成分 |  |  |  |  |
| l | 习用语 |  |  |  | 尚未成为成语的习用语，具有“临时性”，取汉字“临”的声母 |
| m | 数量词 |  |  |  | 取英语数词numeral第三个字母 |
| n | 名词 |  |  |  | 取英语名词noun首字母 |
|  |  |  | nx | 非汉字字符串 |  |
| o | 拟声词 |  |  |  | 取英语拟声词onomatopoeia首字母 |
| p | 介词 |  |  |  | 取英语介词preposition首字母 |
| q | 量词 |  |  |  | 取英语quantity首字母 |
| r | 代词 |  |  |  | 取英语代词pronoun第二个字母 |
| s | 处所词 |  |  |  | 取英语space首字母 |
| t | 时间词 |  |  |  | 取英语time首字母 |
| u | 助词 |  |  |  | 取英语auxiliary第二个字母 |
| v | 动词 |  |  |  | 取英语verb首字母 |
| w | 标点符号 |  |  |  |  |
| x | 非语素字 |  |  |  | 非语素字只是一个符号，字母x通常用于代表未知数、符号 |
| y | 语气词 |  |  |  | 取汉字“语”的声母 |
| z | 状态词 |  |  |  | 取汉字“状”声母的第一个字母 |

1. **标注规范**

①一个词如果在《现代汉语语法信息词典》中已属于某一个或若干个词类，标注时不轻易增加词性。如“训练”、“强调”在《语法信息词典》中只属于动词，标注时切勿仅根据其在当前句子的功能就将它们改为名词或副词，可以标注为名动词vn或副动词vd。

②当《语法信息词典》给某个词确定的词性确实不对或不完备时，当然也要订正或补充。例如，“行政”这个词，在《语法信息词典》中只有名词词性n，但在《人民日报》语料中，诸如“依法行政”这样的用法并不少见，“依法行政”应该加工为“依法/d 行政/v”, 那么，“行政”应当兼属动词v。当然，《语法信息词典》并不一定要补全在语料库中出现的所有词性，也要考虑频度等其他因素。

**词性标注及示例**

1. **形容词（a）**

表示性质、状态，在句中主要充当谓语、定语、状语和补语。如：

大 小 白 红 香 甜 美丽 贫困 富裕

标注示例：

小/a 床/n；

这/r 花/n 很/d 红/a；

繁荣/a 的/u 景象/n 【形容词作定语】；

放/v 明白/a 一些/m 【形容词作补语】

* **形容词兼作动词的情况：a → v**

若该形容词在句子中带了**真宾语**，则标为v。

标注示例：

他/r 跟/p 她/r 没/d 红/v 过/u 脸/n 。/w

繁荣/v 市场/n；端正/v 态度/n

1. **区别词（b）**

表示事物的区别性特征，在句子中只能作定语修饰名词或跟助词“的”组成“的”字结构。如：

男 女 公 母 雌 雄 微型 国产 军用

（1）区别词如果为切分单位，则标为词性b。

标注示例：

女/b 司机/n；金/b 手镯/n

慢性/b 胃炎/n；古/b 钱币/n；横断/b 山脉/n

（2）单音节区别词如果和单音节名词或名语素组合，作为一个切分单位，则标为词性n。

标注示例：

雄鸡/n；女魔/n；古币/n

（3）少数单音节区别词后接双音节词常看做一个复合词（特别是职位）

标注示例：

总书记/n；副主任/n；副教授/n；

总工程师/n；副总参谋长/n

* **区别词兼作副词的情况**

① 若该区别词作状语，则为副词。

标注示例：

我们/r 会/v 共同/d 进步/v

自动/d 取消/v 订单/n

② 若该区别词作定语, 或与“的”组成“的”字结构，则为区别词。

标注示例：

共同/b 目标/n 是/v 完成/v 这/r 项/q 任务/n

这个/r 玩具/n 是/v 自动/b 的/u

1. **连词（c）**

连接词、短语或句子，表示两者之间所具有的某种关系。如：

和 同 与 及 并 或 并且 而且 或者 因为 所 然后 经过 总而言之 由此可见 综上所述

标注示例：

不顾/v 劝告/v 而/c 执意/d 闹事者/n

* **连词兼作副词的情况**

这类兼类词，主要有“不过”、“尽管”、“但”、“可”等。一般来说，若该词在句子中修饰谓语（形容词、动词）则为副词。若此词主要连接句子和子句，表示子句之间转折、让步等语义组合关系，则为连词。如：

① 不过

标注示例：

我/r受/v 了/u 点/q 伤/g ，/w 不过/c 不/d 要紧/a 。/w 【连接句子和子句】

他/r 不过/d 随便/d 谈谈/v 。/w 【修饰谓语】

② 但

标注示例：

雨/n 停/v 了/y ，/w 但/c 地上/s 还/d 很/d 湿/a 。/w

但/d 见/v 门上/s 贴/v 着/u 一/m 副/q 对联/n 。/w

③ 可

标注示例：

大家/r虽然/c 累/a ，/w 可/c 都/d 很/d 愉快/a 。/w

她/r 待/v 我/r 可/d 好/a 了/y 。/w

④ 尽管

标注示例：

尽管/c 天/n 下/v 着/u 雨/n ，/w 他/r 还是/d 出发/v 了/y 。/w

你/r 尽管/d 说/v ，/w 别/d 怕/v 。/w

1. **副词（d）**

修饰或限制动词和形容词，表示范围、程度等，在句子中作状语。如：

都 只 就 仅 很 再三 屡次 将 不 却 总共 正在 常 重新 曾经 竟然

居然 常常 仅仅 原原本本 确确实实 别 老

标注示例：

安/v 不/d 上/v

好好/d 干/v 吧/y

久久/d 没/d 说话/v

轻轻/d吊/v起/v 又/d 轻轻/d 地/u 放下/v

1. **叹词（e）**

表示感叹、呼唤或应答，可独立成句或在句中充当独立成分。如：

啊 嗯 唉 哎 哼 哦 哎哟 哎 嘿 嚯 喂

1. **方位词（f）**

表示位置的相对方向。如：

上 下 左 右 前 后 里 外 中 东 西 南 北 前边 左面 里头 中间 外

标注示例：

近/a 20/m 年/q 来/f

南/f 大街/n‘

1. **语素（g）**

指汉字字符集中一般不单独使用的汉字。语素和非语素字一般不作为切分单位，但是：

1. 某些双音节离合词分开使用，其中一个是语素，可将它标注为语素。

标注示例：

理/v 了/u 一/m 次/q 发/g；

洗/v 了/u一个/m 舒舒服服/z 的/u 澡/g

1. 单字名词/名词性语素 + 单纯方位词

① “单字名词+ 单字方位词”的组合，切分为两个单位，分别标注。

标注示例：

饭/n 前/f；树/n 上/f；包/n 里/f；床/n 下/f

② “名语素字+单字方位词”的结构，不切分，合为一个处所词或时间词，整体标注。

标注示例：

桌/g 上/f → 桌上/s；

午/g 后/f → 午后/t；

身/g 上/f → 身上/s；

胸/g 前/f → 胸前/s

③ “省、市、县、乡、村、部、局、处、团、营、连、院、系、班”等名词后接有“里、上”等方位词，仍有组织、机构的意义，作为一个切分单位，标为名词，不单独切分和标注语素。

标注示例：

部里/n；县里/n；村里/n；系里/n；班上/n

1. **前接成分（h）**

词根前面的附加构词成分。如：

阿 老 初 第 小 大 非 超 无 过

是否标注取决于前接成分有没有和后面的语素或词切分开。

标注示例：

阿哥/n；小王/n 【未切开，不标注】；

非/h 国家/n 工作/v 人员/n 【切开，标注】；

非/h 本市/r注册/v 车辆/n

1. **成语（i）**

标注示例：

成千上万/i 的/u 群众/n；

近水楼台先得月/i；

一年之计在于春/i；

不管三七二十一/i；

众人拾柴火焰高/i；

挂羊头/i ，/w 卖狗肉/i

1. **简缩略语（j）**

专有名词或常用语的简缩形式。如：

人大 五四 奥运 调研 离退休 短平快 高精尖 中 美 日

教科文 中共 三好 省直 老弱病残 农工牧副渔业 中西方

标注示例：

京/j 津/j 唐/j 高速公路/n；

中/j 、/w 美/j 、/w 日/j；

农/j 、/w 林/j 、/w 牧/j 、/w 渔/j 等/u 副业/n；

中/j 美 /j 跨国公司/n；

国（边）境/j

1. **后接成分（k）**

词根后面的附加构词成分。如：

子 儿 头 化 们 式 性 者

是否标注取决于有没有和前面的语素或词切分开。

标注示例：

花儿/n；画儿/n；

人们/n；哥儿们/n；爷儿们/n；

经过/p 苦苦/d 追求/v 而/c 获得/v 幸福者/n；

不顾/v 劝告/v 而/c 执意/v 闹事者/n

1. **习用语（l）**

一种相沿习用的定型短语。如：

跑龙套 打官腔 吃老本 说得过去

1. **数词（m）**

表示数目和次序，包括基数词、序数词、小数、分数、百分数、约数。如：

0 一 半 百 千 百万 一百零八 第一 第十八 19

一个个 一阵阵 一团团 一辆辆 许许多多 多多少少

标注示例：

一百二十三/m；120万/m；

第一/m；20% /m；

几十/m 人/n；

约/d 一百/m 多/m 万/m；

仅/d 一百/m 个/q；

九/m 、/w 十点钟/t；

一/m 慢/a 二/m 看/v 三/m 通过/v

1. **名词（n）**

表示人或事物的名称，在句子中主要充当主语和宾语。

* **专有名词**

1. 人名

标注示例：

张仁伟/n；江/n 总书记/n；老张/n; 鲁迅/n；小林多喜二/n；卡尔·马克思/n

1. 地名

标注示例：

安徽/n；中国/n；四川省/n；东京都/n；喜马拉雅山/n；中关村/n；学院路/n

1. 长地名

标注示例：

宁夏/n 回族/n 自治区/n；

华盛顿/n 特区/n；

台湾/n 海峡/n；

京东/n 大/a 峡谷/n；

北京市/n 海淀区/n 海淀镇/n 南/f 大街/n 蒋家/n 胡同/n 24 /m 号/q

1. 团体机构

标注示例：

联合国/n；

中共中央/n

1. 短语型团体机构

标注示例：

中国/n 计算机/n 学会/n；香港/n 钟表业/n 总会/n；

全国/n 信息/n 技术/n 标准化/v 委员会/n；

全国/n 总/b 工会/n；

联合国/n 教科文/j 组织/n；

安徽/n 人大/j 常委会/j 办公室/n

1. 名词+单语素

标注示例：

满族/n；汉语/n；哈萨克人/n；津浦路/n

1. 事件/运动

标注示例：

卢沟桥/n 事件/n；

五四/t 运动/n

* **普通名词**

1. +多音节名词

标注示例：

欧洲/n 语言/n；

京九/j 铁路/n；

宋朝/t 古董/n；

康师傅/n 方便面/n

1. 以序号命名的名称

标注示例：

2/m 号/q 国道/n；

十一/m 届/q 三中全会/j

1. 报纸书名

标注示例：

《宁波/n 日报/n》；

红楼梦/n

1. 活动等

标注示例：

第三/m 次/q 横田/n 基地/n 噪音/n 诉讼/v

* **名词兼作量词的情况：n → q**

1. 汉语中的一些名词（主要是单音节名词）可以兼作量词，对于这些词，依据上下文来确定句子中的词的词性。比如：数词＋ n-q ＋ n，取 q。

标注示例：

一/m 车/q 煤/n；

三/m 桶/q 水/n

1. 另外，汉语中有一部分名词临时作量词且只能前接数词“一”，对于这种情况，也是应该把它标为量词q。

标注示例：

做/v 了/u 一/m 桌子/q 菜/n；

生/v 了/u 一/m 肚子/q 气/n

1. “这”,“那”,“每”等指示代词＋n-q＋n, 取q。

标注示例：

这/r 床/q 被子/n

这/r 门/q 功课/n

1. **非汉字字符串（nx）**

关于非汉字字符串的标注：

1. 已经约定俗成的或科学技术中已经通用的符号保持原有的意义，根据其原有意义决定相应的标记。

① 阿拉伯数字

标注示例：

121/m号/q房间/n；

2000年/t 8月/t 15日/t

② 单独的罗马数字

标注示例：

Ⅱ/m；

ⅠⅩ/m；

③ 英文字母（或字母组合）代表常用的度量单位，比如A代表“安培”；V代表“伏特”；

W，“瓦特”；m，米；kg，千克；等等。

标注示例：

然后/c指针/n回指/v在/p 1.5/m A/q 处/n 。/w

1. 其他英文字母（或字母组合或语句）一律标注为nx。

标注示例：

世界杯/n 足球赛/n A/nx 组/n 的/u 两/m 场/q 比赛/n；

A/nx 公司/n，B/nx 先生/n，X/nx 君/g；

C/nx 是/v 光速/n；

Windows98/nx；

PentiumIV/nx；

I/nx LOVE/nx THIS /nx GAME/nx

1. 其他西文（希腊文、俄文等）的处理同英文。
2. 日文假名处理同英文。日文中的汉字处理同中文，但不能保证切分的正确性。
3. **拟声词（o）**

模拟自然界事物的某种声音，不能单独成句。如：

砰 滴答 扑通 咕咚 叮叮当当 哗啦哗啦 呼呼 咔嚓 隆隆 咕嘟 汩汩 嘎嘎

1. **介词（p）**

引介名词性成分，不单独充当句子成分。如：

为 在 把 让 比 给 被 从 向 对 凭 按照 对于 为了 自从 关于 根据 依照 凭

标注示例：

经过/p 苦苦/d 追求/v 而/c 获得/v 幸福者/n

* **介词兼作动词的情况：p → v**

这类词主要有“在”，“到”，“比”，“朝”，“跟”，“给”等，它们的区分主要依据以下方法：

① 从词的语法功能与分布考虑，若该词（包括带“着、了、过”的情况）单说或单独作谓语，则为动词。

标注示例：

“你/r 爸爸/n 在/v 不/d 在/v ？/w” “在/v 。/w” 【单独作谓语】

北京/n 到/v了/y 【带“了”】；

新加坡/n 我/r 到/v 过/u 【带“过”】；

别/d 老/d 跟/v 着/u 【带“着”】；

咱们/r 比/v 一/m 比/v 【单独作谓语】

② 对“p-v＋其他成分”的结构，若单说或单独作谓语，则其中的p-v为动词；若不是单说也不是单独作谓语，而是作状语或补语，则其中的p-v为介词。如：

作动词时：

他/r 不/d 在/v 教室/n 【“在”作谓语】；

他/r 在/v 不/d 在/v 家/n ？/w 在/v 。/w

列车/n 已/d 到/v 了/u 北京/n 。/w 【“到”作谓语】

葵花/n 向/v 太阳/n 【“向”作谓语】

作介词时：

他/r 在/p 教室/n 自习/v 。/w 【“在”作状语】

老王/n 到/p 北京/n 出差/v 去/v 了/u 。/w 【“到”作状语】

* **介词兼作连词的情况**

常见的词有“和”、“跟”、“同”、“与”，这些词的词类排歧主要依据下列原则：在句子中，如果这些词的前后成分不能互换位置或者在这些词的前面可以加修饰成分，则这些词为介词；如果这些词的前后成分可以互换位置即互换位置后句子的意思基本不变并且在这些词的前面不能有修饰成分，则这些词为连词。

标注示例：

我/r 跟/c 他/r 都/d 是/v 大学生/n 【可以互换位置】；

你/r 别/d 跟/p 他/r 跑/v 【不可以互换位置】；

我/r 跟/p 他/r 请教/v 问题/n 【不可以互换位置】

我/r （已经/d ） 和/p 他/r 见面/v 了/y 。/w 【前面可以添加修饰成分】

我/r 和/c 他/r （已经/d ） 见面/v 了/y 。/w 【前面不可以添加修饰成分】

1. **量词（q）**

表示人、事物或动作的单位。如：

个 条 片 匹 辆 尺 斤 两 吨 支 回 次 遍

千瓦时 千 号 届 年 元 天 岁 公斤 张张 个个

或是 已被录入《语法信息词典》中的少数数量词。如：

一个 一些 百年 半天 一会儿 一整套

标注示例：

三/m 个/q；一/m 盒/q 点心/n

1. **代词（r）**

起替代和复指作用。如：

我 你 他 这 那 谁 本 每 各 诸 某 该 全 此 我们 你们 他们 那些 那样 那里

这个 这么 这边 那个 大家 自己 什么 哪里 怎么 怎么样 些 多少 怎样 为什么

（1）单音节词作代词，如果不与后面的名词（后接词也为单音节）切开，则整体合为代词。

标注示例：

本报/r；每人/r；本社/r

（2）当后接的名词有2个以上音节时，先切分，再进行词性标注。

标注示例：

本/r 地区/n；

各/r 部门/n；

贵/r 出版社/n

1. **处所词（s）**

表示处所。如：

空中 高处 隔壁 门口 附近 边疆 一旁 野外

标注示例：

雨/n 停/v 了/y ，/w 但/c 地上/s 还/d 很/d 湿/a 。/w

但/d 见/v 门上/s 贴/v 着/u 一/m 副/q 对联/n 。/w

1. **时间词（t）**

包括一般所说的时量词。如：

年 月 日 分 秒 现在 过去 昨天 去年 将来 宋朝

（1）“牛年”“虎年”等不切分且标为t，“甲午年”“庚子”“戊戌”等也标注为t。

标注示例：

甲午/t 战争/n

3月/t 10日/t 下午/t 2时/t 19分/t

* **注意区分数量词与时间词**

例如：“78年”指“1978年”时应标注为 “78年/t”，

当指数量“七十八年”时应切分标注为“78/m 年/q”。

再如：两/m 个/q 月/n

三/m 天/q 时间/n

同样：当“8日”指一个月当中的第八天时为时间词，不予切分，标注为 “8日/t”，

若表示8天时段，则要分开，标注为 “8/m 日/q”。

（2）如果数字后面没有表示时间的时量词（年月日等），标为数词m。

标注示例：

1998 /m 中文/n 信息/n 处理/n 国际/n 会议/n

1. **助词（u）**

附着在词、短语、句子后面表示某种附加意义。如：

的 地 得 了 着 过 等 似的 一样

标注示例：

出/v 了/u 一/m 次/q 差/g 。/w

安/v 得/u 上/v

1. **动词（v）**

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

* **动词述补结构的标注**：

① 拆分的情况下标注示例：

走/v 到/v；

调 /v 好 /a；

走/v 得/u 到/v；

走/v 不/d 到/v

② 不拆分的情况下标注示例：

形成/v；

来得及/v；

对不起/v

* **动词兼作名词的情况：v → n**

当该词表示一种动作时，后面带真宾语，则标为v；当它指称人或物时，则标为n。

标注示例：

编辑/v 科技/n 文献/n 【带真宾语】；

她/r是/v 责任/n 编辑/n 【指称人或物】；

要/v 锁/v 上/v门/n【带真宾语】；

忘/v 了/u买/v 一/m 把/q锁/n 【指称人或物】；

及时/a 报告/v 首长/n 【带真宾语】；

一/m 份/q 重要/a 报告/n 【指称人或物】

1. **标点符号（w）**

标注示例：

北京/n - /w 西雅图/n 航线/n

1. **非语素字（x）**

非语素字只是一个符号，字母x通常用于代表未知数、符号。

标注示例：

鹧鸪/n 的/u 鹧/x 有/v 什么/r 意思/n 吗/y ？/m

1. **语气词（y）**

吗 吧 了 啊 啦 唉 呢 哇 呀 哦 哈 哟 么 焉 矣 乎 耶等词。

标注示例：

好好/d 干/v 吧/y

北京/n 到/v 了/y

1. **状态词（z）**

**【注意这一部分词语在分词文档里下属形容词，在标注词性时标注为z】**

常见状态词形态有：

“的” “地”；双音节形容词的重叠形式ABB；二字形容词的重叠形式AABB；

有两个意义相反的单字形容词并列而成的名词再重叠所得到的重叠形式“AABB”；

双音节形容词的重叠形式“A里AB”；部分A不AB形式的词语

标注示例：

甜甜的/z 点心/n；

削/v 得/u 尖尖的/z；

雪白/z 雪白/z；

圆圆地/z 坐/v 一/m 圈/q

孤单单/z；亮堂堂/z；

高高兴兴/z；舒舒服服/z；

大大小小/z；高高低低/z；

马里马虎/z，糊里糊涂/z，慌里慌张/z；

容不容易/z；漂不漂亮/z；