# 分词标注文档

本文档主要参考《The segmentation guidelines for the Penn Chinese treebank 3.0》

## 测试是否为词粗略规则

* 1. 粘着语素

通常情况下，粘着语素­­­­会和周围的语素连接组成一个词。粘着语素是不可能单独成句的语素。

* 1. 复杂的内部结构

一串有着复杂内部结构的文字，通常需要被拆分

* 1. 组合性

如果X-Y表示的意思是非组合性的，那么X-Y 很可能是一个词。

* 1. 插入

如果一个其他的语素可以插入到X和Y中间，那么X-Y 很可能不是一个词。

* 1. 短语替代

如果一个语素不能被同类型的短语替代，那么这个语素是一个词的一部分。

* 1. 音节数

有时可以考虑X-Y 这两部分是否为单音节(可简单理解为是否是一个字)。

## 切分细则

假设已经将句子分切分成了稍大些的块，对每块可根据细则来考虑是否需要继续切分。

黏着语素一般不切分。

从字数考虑，两个字的组合较宽地看作是一个切分单位，三个字的较严，四个字以上的如果不是成语、习用语，一般不看作是一个切分单位。

《现代汉语语法信息词典》数据库中字段“词语”（包括词、词组、语素、前接成分、后接成分、习用语、简称、标点符号等）都已归类，且带有词性标记。

一般已经归入词典内的都是切分单位，对真实文本的覆盖率很高，但还是会存在例外。比如：

* 5个字以上的成语、习用语是切分单位，但未被收录进词典中；
* 像“一百二十八”、“五分之三”、“百分之九”、“1998年”、“10月30日”这样的数词和时间词实际上是无限多的，《语法信息词典》不可能全收，只可能收少量的构成成分，它们都作为一个切分单位；
* 像“分之”、“百分之”作为助数词收入了语法信息词典，但它们并不是切分单位，不与前接和后接的数字切开；
* 语法信息词典中包含的前接成分、后接成分、语素、非语素字都不是切分单位，尽管当它们不能与前后成分组合时也会孤立地出现在切分序列中；
* 未登录词。

### 普通名词

下面提到的名词 并不代表其在汉字数量上大于一个字，比如 『鱼』也是一个名词。 后面章节提到的动词等其他词性同等类比。

#### 亲属名

把亲属名看做一个词，均不切分。

比如：三叔，表叔，大姑父

#### 数词 + 名词

如果一个度量词 (如：米、件、名等) 可以插入到 数词 – 名词 之间，并且不改变原来表达的意思，那么我们把 数词 和 名词 切分，否则不切分。

切分例子：一 学生（中间添加”名”不改变意思，所以切分）。

不切分例子：半边 （半边也已登陆词表）

#### 区别词 + 名词

如果 **区别词** 或者 名词 是单音节的，并且其中一个语素是粘着的，则不做切分；否则将区别词和名词切分。

不切分例子：雄鸡，雌象，女魔，古币

切分例子：女 司机 金 手镯 慢性 胃炎 古 钱币

共同 利益 （后接名词为两个音节，所以切分）

特殊：少数单音节区别词后接双音节词但看做是一个复合词，不切分，特别是职位。

不切分例子：总书记 副主任 副教授 总工程师 副总参谋长

#### 代词 + 名词

如果 **代词** 和 名词 都是单音节的，并且名词是粘着的，则不切分；否则将代词 和 名词切分。

当后接的名词有2个以上音节时，切分。

不切分例子：我校 本报 每人 本社 本校

切分例子：我 单位 本 地区 各 部门 贵 出版社

#### 前缀/修饰成分（形容词）+ 名词

**X** + 名词的形式。X可以是个**前缀**，或者是个**修饰成分**。

根据 X 的形式决定是否切分：

如果X 是一个**前缀**，那么不切分；如果X是一个**修饰成分**，那么切分。

修饰成分和前缀的区别在于：前缀是粘着的。可以参考词表，或者使用一个

名词短语来替换原词中的名词，如果可以被替换，X就是修饰成分，否则为

前缀。

* 如果X 是**前缀**，则不切分。

前缀举例：阿，非，未（此处只针对“非”和“未”等作前缀的情况，“非”和“未”等还可以作动词/副词）

不切分例子：阿爸 非暴力（ “非”此时是前缀，词性为Ng）

类型 未定稿 （此时“未”为前缀，词性为Ng）

特殊：若“非”等所管辖的范围超过一个词，则仍然切分开。

切分举例：非 国家 工作 人员；非 本市 注册 车辆

* 如果X 是**修饰性成分**，并且其和名词都是单音节的，那么不切分，否则切分。修饰成分举例：原，前

不切分举例：女人 （女 是个 非谓语形容词）

切分举例：前 民主德国

非谓语

定义: 如果一个词 只能在『是 … 的』的帮助下，作为主语后面的谓语，那么这个词就是非谓语形容词。

* 如果X是一个**形容词**

如果X 和 名词 都是粘着的，或者 X + 名词 的含义是非组合性的，那么不切分。在不清楚的情况下，如果 X 和 名词 都是单音节的，那么都不切分。

不切分举例：小媳妇，大洲，大海

切分举例：厚 书（书不是粘着名词，自己可以成词） 小 妙招

#### 方位词 + 名词

如果都是单音节的，那么就不切分。

不切分举例：前院，前天，左肩

切分举例： 左 肩膀（肩膀 这个名词不是单音节，所以切分）

#### 名词 + 方位词

如果是 名词 + 方位词1 + 方位词2 的格式，如果 方位词1 和 方位词 2 代表相反的方向，则视作 名词 + 方位词 格式。

如果以下三点都满足，则不切分：

1. 名词和方位词都是单音节
2. 在语境中名词是粘着的，或者无指的（non-referential）
3. 在语境中名词不能被 限定词-度量词 修饰。
4. “省、市、县、乡、村、部、局、处、团、营、连、院、系、班”等名词后“里、上”等方位词，仍有组织、机构的意义。

否则切分。

不切分举例：室内，台下，跟前，境外，境内外，海外，背后，天下，国内

午后，赛前，午后；身上，胸前，部里；县里；村里；班上

切分举例：中午 以后；饭 前；树 上；包 里；床 下

#### 名词1 + 名词2： 名词1 修饰 名词2

如果名词1 有一或两个音节，名词2只有一个音节，则不切分。

如果一个不超过两个汉字的名词，后面跟随着多个接尾词，则整个字符串，

当做一个词不切分，比如：物理学家。（”学”和”家”都是接尾词）

其他情况下，两个名词需要切分。

不切分举例：北京市；研究室；发展史；始祖鸟；残疾人；

清晰度；紧迫感；大奖赛；太阳系

切分举例：北京 大学； 玩具 工厂；合作 领域； 史学 研究

#### 动词 + 名词

这里动词 不包含 『谓语形容词』 和 『有』『是』等动词。

谓语形容词参考1.5章节。

动词参考9.7章节。

如果动词是修饰名词的，则不切分。

不切分举例： 烤肉；炒菜；证明信；讨论会

### 专有名词

#### 人名

人名看做一个词，但有空格时切分。

不切分举例：张胜利；

切分举例：John Smith 卡尔 马克思

#### 带前后缀的人名

不切分，视为一个单词。

不切分举例：老张， 张老

对人的简称、尊称如果是两个字，也不切分。

不切分举例：大李 郭老 小郝 陈总

#### 人名 + 称谓

切分，当做两个词

切分举例：张 教授； 张 李 两 位 教授

#### 地名/团体/机构/组织/事件/运动等

如果是由 名词1 + 名词2 构成，如果 名词2 是个普通名词，并且名词2 是单音节，则不切分；否则切分。

不切分举例：北京市；黄河；沙市；黑龙江省

切分举例：北京 大学；北京 第一 服装厂；美国 国会

（1）国名不论长短，作为一个切分单位。

不切分举例：中国 美国 日本国 阿富汗 巴勒斯坦国

（2）地名后有“省”、“市”、“县”、“区”、“乡”、“镇”、“村”、“旗”、“州”、“都”、

“府”等单字的现代行政区划名称时，不切分开，作为一个切分单位。

不切分举例：四川省 天津市 景德镇市 华盛顿州

俄亥俄州 东京都 大阪府 开封府

如果地名后的行政区划有两个以上的汉字，则将地名同行政区划名称切开。

切分举例：芜湖 专区 内蒙古 自治区 宁夏 回族 自治区

地名后有表示地形地貌的一个字的普通名词，如“江、河、山、洋、海、岛、

峰、湖”等，不切分。

不切分举例：鸭绿江 亚马逊河 喜马拉雅山

如果地名后接的表示地形地貌的普通名词有两个以上汉字，切分。

切分举例：台湾 海峡 华北 平原 帕米尔 高原

地名后有表示自然区划的一个字的普通名词，如“街，路，道，巷，里，町，

庄，村，弄，堡”等，不切分。

不切分举例：中关村 长安街 学院路 景德镇 北海道

如果地名后接的表示自然区划的普通名词有两个以上汉字，切分。

切分举例：蒋家 胡同 陶然亭 公园

（3）团体/机构/组织等：

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，切分。

切分举例：中国 计算机 学会 北京 图书馆 富士通 株式会社

（4）事件/运动等：

事件、运动等一般为短语型，按短语型专有名词处理，切分。

切分举例：卢沟桥 事件；明治 维新

#### 专有名词 + 专有名词：无连接词协调

看做两个词，切分。

切分举例： 中 美； 东 新 澳；欧 美；日 韩

* 1. **以序号命名的名称一般不认为是专有名词**

切分。

切分举例：2 号 国道；十一 届 三中全会

* 1. **书/报纸/杂志/文档等**

书、报、杂志、文档、报告、协议、合同等的名称通常有书名号加以标识，

不作为专有名词。由于这些名字往往较长，名字本身按常规处理。

切分举例：《宁波 日报》；杜甫 诗选；陆俭明 自选集

少数收入词典的书名、报刊名等专有名称，则不切分。

不切分举例：红楼梦；人民日报；儒林外史

* 1. **无法分辨是人名/地名/机构名**

切分。

切分举例：巴黎 贝尔希 体育馆

* 1. **大型活动/商标**

一般的命名活动常用引号表示，不看作专有名称，进行切分。

切分举例：迎 香港 回归 京九 植绿护绿 活动

商标（包括专名及后接的“牌”、“型”等）是专指的，，但其后所接的商品仍是普

通名词。切分。

切分举例：康师傅 方便面 中华牌 香烟 联想 电脑

* 1. **食谱**

食谱不看作专有名词，食谱上的菜名等通常也是短语型的，如果拆开了，意思差别很远，则不切分，否则切分。

不切分举例：宫保肉丁；木樨肉；松鼠鳜鱼

切分举例：鸡蛋 汤；芝麻 饼

* 1. **专有名词+单语素**

专有名称后接单音节的语素，如表示民族的“族”、表示语言的“语”，表示文字的“文”， 专有名称后接单音节的名词，如表示人种的“人”、表示奖项的“奖”，不切分。

不切分举例：满族 汉语 中文 哈萨克人 诺贝尔奖

* 1. **专有名词+多音节名词**

专有名称后接多音节的名词，如“语言”、“文学”、“文化”、“方式”、“精神”等,

切分。

切分举例：欧洲 语言 法国 文学 贝多芬 交响乐 雷锋 精神

### 其他成分+名词

JJ 需要与它修饰的名词，以及 度量词 切分开。 比如： 三 大 杯 水

当JJ修饰名词时，JJ 可以是形容词，非谓语形容词，或则短语动词。 大部

分短语动词 都由 X-Y 两部分组成，X-Y 都是单音节的，并且X和Y都是

一个词的缩写形式。下面就来举例说明一下X-Y 短语。

#### 动宾结构 + 名词 【动宾结构见9.7】

比如：发稿（发稿 时间）；

驻华 （驻华 使馆）；获奖（获奖 学者）

#### 副词 + 谓语形容词 修饰名词

如 ： 最新（最新 消息）； 超大（超大 规模 集成 电路）；

**（注意：最新 是形容词，超大 是区别词）**

常见的副词有：最，超，较

特殊：最+形容词 作形容词

1. 如果是“最+形容词+的+名词”的形式，最与形容词切开，例如：

最 强 的 人类

1. 如果是“最+形容词+名词”的形式，最与形容词不切开，例如：

最强 人类

* 1. **谓语形容词 + 名词 修饰名词**

如： 高层（高层 人士）； 高速（高速 公路）；大幅（大幅 标语）

#### 数量词 + 名词 修饰名词

如： 两 国 （两 国 关系）； 多 国（多 国 部队）

* 1. **介词 + 名词/方位词 修饰名词**

如： 对外 （对外 政策）

#### 3.6 其他 修饰名词

如：关贸（关贸 总协定）； 年均（年均 增长率）；

历届（历届 世界 体操 大赛）；上述（上述 三 国）；

有关（有关 方面）

### 4. 时间词

年、月、日、小时、分钟 等等，表示时间 几点  几月 几日，都要被看成一

个单独的词，不切分。

切分举例：1998年 3月 21日；五点钟；初一； 去年

#### 数词 + 名词 表示时间

如果 数词 + 名词 指的是一个时间的名称（具体的时间）那么不拆分，看做一个词。如果是代表时间的多少，则拆分。

不切分举例：1998年； 5点钟； 90年代

切分举例：3 年； 3 个 月

### 5. 方位词

方位词参考 **1.6；1.7；以及6.1**。

方位词通常为一个或两个音节：

单音节方位词举例： 内， 后

双音节方位词举例： 之间，以来， 以后， 左右, 前后

### 6. 代词

代词自身看做一个词，不切分

举例： 他们；他自己；自己；多少；怎样；哪里；什么；

每年；怎么样；怎么办；哪般 哪边 哪个 哪儿 哪会儿 哪会子 哪里 哪门子 哪些 哪样 每次 每个 每年 每月 每日

这边 这儿 这个 这会儿 这里 这么 这么点儿 这么些

这么样 这些 这样

那边 那儿 那会儿 那里 那么 那么点儿

那时 那位 那样 那种 怎么办 怎么样 怎么着

#### 6.1 代词 + 方位词 都是单音节

如果代词和方位词都是单音节，则不切分，视作一个词。

不切分举例：此间；此前；其中；

切分举例：这 以后

#### 6.2 代词 + 方位词 在句中作副词

如果 代词+方位词，在句中作用是副词的话，则不切分。

举例：此外

### 7. 数词

1. 纯数字： 一亿三千万；30.1；123,456；35.6%；30万

基数、序数、小数、分数、百分数一律不予切分。

不切分举例：第一 第十五 三分之二

2. 约数： 三四十 岁

3. 数字+X+数字： X 是个像【余、点】类似的词素。

如：三十 几 亿； 好 几 个；两 点 多

4. 数字+X： X 是像【余，来】类似的词素. 如： 三十 来 个；四千一百 余

**【注意】**

1. 如果有顿号，则切分。

切分举例：五 、 六 年

1. 已被《语法信息词典》收录的，不切分。

不切分举例：**一个**；一些；百年；半天；一会儿；一整套

1. 表序关系的 数词 + 名词 结构，切分。

切分举例：一 营；三 连

1. 数量结构的“ABB”“AABB”形式，不予切分。

不切分举例：一个个 一阵阵 一团团 许许多多

1. 些、一些、点儿、一点儿，不切分。

不切分举例：懂 一些 快 一点儿

需要区分的情况：

当“一点”用作数词，表程度时，“一点”不切分；

当“一点”作名词时，“一点”需要切分。

切分举例：有 一 点 需要 注意 的

### 8. 量词

把度量词，重复或者混合的度量词，都看做一个词不切分。

比如：杯；杯杯；架次； 分钟

数量词组切分为 数词 + 量词。

切分举例：三 个；10 公斤

### 9. 动词

动词包括五类：

谓语形容词：红

系动词：为，是

形式动词： 有

其他动词： 走

#### 9.1 重复动词1：AA, ABAB, AABB, AAB, ABB, ABAC

当做一个词，不切分。

* AA， A是一个动词：

看看；红红

* ABAB： AB 是一个动词

如： 研究研究； 雪白雪白

* AABB, AB 是个动词

如： 来来往往； 高高兴兴

大多数情况下 AA 和 BB 并不是个动词

* AAB（排除 AA-看，2.10.2）

如：蒙蒙亮

* ABB

如： 红彤彤， 绿油油

* ABAC

如：马里马虎，有条有理，一清二楚

* ABAB 式的切开

如：研究 研究 高兴 高兴 很多 很多 雪白 雪白 一个 一个

#### 9.2 重复动词2：AA-看，A一A，A了一A， A了A

需要切分开。

* AA-看

如： 说说 看

* A一A

如：想 一 想

* A了A

如：想 了 想

* A了一A

如：想 了 一 想

注意： 动词 + 数词 + 测度， 需要切分，比如： 看 一 眼

#### 9.3 A-not-A

这里not 代表 表示否定的词，需要切分。

比如： 来 没 来；

不 写；看 不 看；未 完成；没 研究

但是如果形成“A不AB”的不完整形式，则不予切分。

如：相不相信 容不容易 漂不漂亮

#### 9.4 形容词 + 动词 【偏正结构动词】

满足以下几点的其中之一，则不切分。

1. 没有词能插入到形容词 和动词 中间。
2. 没有这个形容词，动词不能做谓语。
3. 形容词 + 动词的 所属的 子语类框架（subcategorization frame），与单独动词 不相同。

否则，切分成两个词。

不切分举例： 胡说；胡来；敬献；尚余；历任； 并列；不畏；乱发

切分举例：已经 采取；早 来；晚 走；老 是

#### 9.5 其他小品词 + 动词

如果去除 其他小品词，动词单独不能作为谓语，则不切分，看做一个词。

举例： 以期；

#### 9.6 名词 + 动词

根据上下文，一些 主语 + 谓语 字符串，可以为一个短语，或者为一个词。

如果一个 修饰谓语形容词的词 可以被插入到主谓之间，并且主语是有指（referential） 的。那么这个字符串要被看做短语，否则看做一个单词。

不切分举例： 头疼 ( 他 让 我 很 头疼)

切分举例：我 头 疼。（头 与疼 之间 可以可以插入 『很』 来修饰，

所以切分。

#### 9.7 动词 + 名词 【动宾结构】

如果动词和名词之间 被 时态标志词（如 了） 隔开，或被修饰 名词的词隔开，则切分成两个词。

如果动词和名词之间是相邻的，按如下规则:

* 如果 动词 + 名词 组合是及物的，并且宾语 只有在 动词 + 名词 组合之后出现，不能单独在当前的名词后出现，则不切分。（比如：投资；出席；关心；为期）
* 如果 动词 和 动词 + 名词 有相似的含义，并且都是不及物的，那么不切分。（比如：睡觉）
* 如果名词是粘着的，那么不切分（比如：游泳；无望；无效；无法；辞职）
* 如果动词 和 名词 都是一个汉字，并且组成的意思是非合成的（non-compositional）,则不切分。（比如：念书；流血）

切分举例：访 华

* 动宾结构的词或词组中间如果插入了其他成分，则切分。

切分举例：吃 两 顿 饭；跳 新疆 舞

未分动宾结构：断尾 扫地

已分动宾结构：无 骨 取 票 拖 地 浇 地

**【述补结构】**

#### 9.8 动词 + 结果词

动词 + 结果词 是指 动词 和 结果词 **都是动词**，并且存在（动词-的-结果词，动词-了-结果词）这种潜在形式。所以 动词 + 结果 包含了结果和方向的**动词组合**。（比如： 看见；走上来），但是不包含如『改善』，『鼓动』的格式。

我们把符合 动词 + 结果词 定义的词，如果是 动词 或 结果词 是大于两个音节，以及 结果词是『完』则将 动词 和结果词进行切分，否则不切分。

不切分举例： 吃掉；看见；擦净；形成；说明

切分举例： 做 完； 认识 到；走 到；撞 上；调 好；坐 稳

#### 9.9 动词-得/不-结果词

如果 动词-结果词 是存在的，那么就切分；否则不切分。

切分举例： 擦 不 干净； 擦 得 净

不切分举例： 吃不了； 买不起

#### 9.10 动词 + 方向词

如果是复合的趋向动词，一般不切分，但是中间有如果插入“得“或”不“，则切分。

不切分举例：出去 进来【复合的趋向动词】

切分举例： 出 得 去 进 不 来

如果是动词+趋向动词，则切分。

切分举例：寄 来 跑 出 去

#### 9.11 动词 + 动态助词

这种情况下都切分。

切分举例： 走 了

#### 9.12 动词 + 得

这种情况下切分。

举例： 走 得

已登陆动词不作切分，不切分举例：

懂得 觉得

#### 9.13 并列动词，无连接词

如果两个动词都是一个汉字，则不切分；否则切分。

不切分举例：修建

切分举例：宣传 鼓动；

#### 9.14 动词 + 动介词

可以看作 动词 + X， X是单音节，并且为介词 或者是动词。

如果动词是单音节，则不切分，否则切分。

如果X 为『给，为，成，作，到，出，自，向，入，以』中的一个，不切分。

不切分举例：送给 交给 赠送给 翻译成 当作 起到 认识到 来自

面向 流入 报以 加以

特殊情况：

* 如果X 是『在，似』中的一个，则无论动词是否为单音节，均切分。

举例：生 在； 坐 在；深 似 海

* 如果X 是 『于』时，如果此时『于』能被『在』代替，则将动词和『于』切分，否则不切分。

切分举例：生 于；建 于；用 于

不切分举例：等于；缘于；大于；小于；利于

#### 9.15 X + 动词

形如 X + 动词 或 动词 + X ，如果X不能修饰其他的动词，或者 没有X的情况下动词不能做谓语，那么就不切分。

举例： 以期 都是

切分举例：

已登陆的不切分动词举例：类似

#### 9.16 动词 + 代词

切分。

切分举例：怎么 做 去 哪里

#### 9.17 动词 + 方位词

切分。

切分举例：健身 前后 喝酒 后

### 10. 副词

副词修饰短语时，首先要与它修饰的短语分隔开。

切分举例：多 高 多 重

副词修饰数字时，需要切分。

切分举例：近 三十 5 分 多 钟

『极大』这个词 在修饰 动词短语时，要把『极大』整体看做一个副词，而不是 副词+谓语性形容词，因为『大』不能单独的修饰后面的动词短语。

#### 10.1 重复副词

当副词重复时，可以看做一个词，不切分。

举例：常常；仅仅 原原本本 确确实实

#### 10.2 限定词 + 度量词/名词

『这样，同机』在修饰一个动词短语 时看做一个词。

#### 10.3 其他

如果某个短语已经表现的是个副词，则不切分。

比如: 进一步 只有

### 11. 介词

介词要与后面的名词短语切分开。

大部分介词都是单音节的。一些常见双音节介词如下：

为了；随着；沿着；本着；鉴于；除了；经过；作为；截止。

当一个动介词 跟在动词后面时，我们要判断这个动介词是不是和动词组成了一个词。 动词+介词参考 9.14.

#### 11.1 介词 + 代词

这时要切分

举例：为 此

### 12. 连词

需要将连词与其连接的短语切分。

举例：和； 与

内部不切分的连词举例：就是

#### 12.1从属连词

从属连词 需要和后面的短语分割开。

『只有』需要结合语境判断：

 只有 … 才 … ，这时只有是从属连词

 他 只 有 三 块 钱。 『只』是副词 『有』是动词。

### 13. 助词

把助词与其连接的短语分割开。

大部分助词都是单音节的，双音节的组词例子是：**的话；似的**

助词“所”和它后面的动词切分。

切分举例：所 想 所 认识

### 14. 感叹词

看做一个词，内部不切分。

比如：哈

### 15. 拟声词

看做一个词，不切分。

如：哈哈； 哗啦啦；

双音节拟声词“ABAB”，切分。

切分举例：哗啦 哗啦

### 16. 标点

当标点是一个词的一部分时不切分，否则切分出来。

不切分举例：123,456; 卡尔·马克斯

### 17. 外语

看做一个词切分出来，除非是其他词的一部分时，不切分。

如：卡拉OK

### 18. 其他

#### 18.1 成语

成语都不切分

如：各有所好，一比高低

#### 18.2 兼语

小于四个音节不切分，大于等于四个音节时按照停顿切分。

不切分举例：进出口； 国内外

切分举例：交响 乐团； 北京 市长

#### 18.3 缩写

小于四个音节不切分，大于等于四个音节时切分

不切分：三好（三好 学生）；教科文（教科文 组织）；七中（七中 全会）

**18.4 不完整字段**

切分。

### 19. 形容词

#### 19.1 重复形容词 AA

单字形容词重叠式AA，有的成词，有的不成词。如果后面没有“的”，就成

词，切分；如果后面有“地”，切分。如果只有加“的”或“地”才成词，不切分。

切分举例：好好 干 吧 久久 没 说话

轻轻 吊 起 又 轻轻 地 放下

不切分举例：甜甜的 点心 削 得 尖尖的

#### 19.2 重复形容词 ABB AABB

双音节形容词的重叠形式ABB, AABB，不切分。

不切分举例：孤零零 亮堂堂 大大小小 高高低低

#### 19.3 两个单字形容词并列

两个单字形容词并列并且改变词性的，不切分。

不切分举例：长短 深浅 大小

形容词并列但是各自保持原有语义的，切分。

切分举例：大 小 尺寸 光荣 伟大

### 20. 区别词

不切分的区别词举例：一级 三级

### 21. 一些语素搭配

#### 21.1着

一些介词是以着结尾的，如：随着。

#### 21.2之

* 之+方位词，当方位词是单音节时，看做一个词。比如：之外；之中。
* 之+数词，看做 的+数词，需要切分。比如： 方法 之 一； 方法 之 三。
* 为简单起见，『之一』在像『中国 是 发展中 国家 之一』这样的句子中时，把它看做一个词，不切分。
* 之 + 名词，看做 的 + 名词。如：少年 之 家

#### 21.3不

X+不 或 不+X, 如果X必须和 不 同时出现，或者X+不 是 非组合的，我们就把X+不，或 不+X看做一个词。

比如： 不到；不足；不便；不久

切分举例：不 雅

#### 21.4是

『特别是』 当做一个词。

#### 21.5些

这些；一些 看做一个词。

#### 21.6有

动词 + 有 通常是个动词，比如 刻有；具有；富有。

『没有』 看做是一个词。

『有 所；仅 有；有 可能』 看做是两个词。

下面是一些需要结合语境理解的情况：

* 有点

如果 『点』能被『一点』替代，这种情况被视为 动词+度量词处理。

如果『有点』在动词短语之前，且能被其他表程度的副词替换，比如『很』，这时看做一个整体，是个副词。

不切分举例： 他 有点 下不了 台。

切分举例：这 本 书 有 点 意思。

* 有的

有时可以看做 限定词，这时当做一个整体，不切分。

切分举例：他 有 的 书 我 也 有

不切分举例：有的 人 已经 走 了

* 有些

切分举例：我 只 有 些 旧书。

不切分举例：他 不 像 有些 人 专门 爱 抬杠。

* 只有

不切分。

不切分举例： 你 只有 学习 才 能 改进 工作

他 只有 10块 钱

#### 21.7在

『正在』看成一个词。

#### 21.8自己

代词 + 自己，视为一个词。

比如：他自己

### 22. 语素和非语素字

某些双音节离合词分开使用，其中一个是语素，切分。

比如：出 过 两 天 差

理 了 一 次 发

洗 了 一个 舒舒服服 的 澡

语素“类”的情况：

如果“类”前的名词是单字，则不将名词和“类”切分；

如果“类”前的名词为两字或两字以上，则不切分。

不切分举例：酒类

切分举例：抽奖类

### 23. 常见搭配

#### 23.1 不切分搭配

* 副词：迄今为止；迄今；进一步；越来越；同机；沿途；即席
* 限定词：这些
* 名词修饰词JJ: 对外，各界
* 方位词：对外；在内
* 其他名词：其中
* 介词：为了
* 动词：来自；面向；流入；迈向；报以；为期；有利于

#### 23.2 切分搭配

* 副词类：并 未
* 连词类：及 其； 而 又
* 限定词类：各 个
* 其他名词： 超大 规模； 我 国
* 时间名词：零点 零一分

#### 23.3 其他

动词-动词：迎上 前去