# HZAU CQPweb 使用手册

## 平台登陆

登录网址: http://aec.hzau.edu.cn

用户名：test 密码：test

## 登陆界面

| | | |
|---|---|---|
| **Journal Articles 2.0**<br>农科专业期刊论文语料库 | | |
| **Articles by Disciplines**<br>全文语料库 | **Article Abstract**<br>摘要子库 | **Article Conclusion**<br>结论子库 |
| **Article Discussion**<br>讨论子库 | **Article Introduction**<br>前言子库 | **Article Method**<br>方法子库 |
| **Article Result**<br>结果子库 | **Articles by Sections**<br>章节子库 | |
| **earner Papers 2.0**<br>汉语母语学习者语料库 | | |
| **Journal Articles** | **Learner Papers** | 扩大库容前的版本<br>不建议使用 |
| **Other Corpora** 其他语料库 | | |
| **BAWE**<br>英国学术英语语料库 | **Dickens**<br>狄更斯作品语料库 | **3in1** |

以下举例中**正体**表示搜索框中的输入内容,箭头表示检索结果,以**斜体**显示。

## 标准检索-简单检索 Simple query

进入**[Standard query]**页面(默认),在检索模式**[Query mode]**下拉菜单中选择简单检索**[Simple query]**,搜索框中输入单词、短语等,点击**[Start Query]**进行检索。

- **[Simple query]**默认情况下为**[ignore case]**,意为检索不区分大小写(例:bath 和 BATH 检索结果相同)；如果选择**[case-sensitive]**,则区分大小写。
- Simple query 的检索方式见命令输入框下面的帮助文件 Simple Query Syntax help,注意帮助文件中的词性赋码已从 CLAWS5 升级为 CLAWS7,但其检索方式仍然可用

## 标准检索-复杂检索

进入**[Standard query]**页面,在**[Query mode]**下拉菜单中选择复杂检索**[CQP Syntax]**,**此时区分大小写**

- 主要查询的属性是 **word(单词)**和 **part of speech (pos, 词性赋码)**,查询时采用**[属性="值"]**的形式

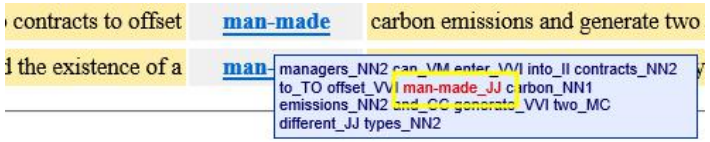
  [word="take"] ➔ 所有包含单词 *take* 的索引

  [pos="NN1"] ➔ 所有可数名词单数
- 最基本的几种 pos **词性赋码**:

| | | | |
|---|---|---|---|
| N.* | 名词 | I.* | 介词 |
| V.* | 动词 | AT.* | 冠词 |
| J.* | 形容词 | VB.* | be 动词 |
| R.* | 副词 | VH.* | 有动词 |
| VVN | 过去分词 | VV.* | 实意动词 |
| VVG | 现在分词 | VM.* | 情态动词 |

- pos 词性赋码的完整列表见附录
- 在检索结果页面,把鼠标放在任意单词上悬停即可查看该词的**词性赋码**

  如搜索"man-made",在结果页面可以看到,man-made 后面的词性赋码为 JJ,即普通形容词




- 一个**单词**或**标点**称为一个**检索单元**
- 每个检索单元间以**中括号[]或空格**作为分隔符

  [word="take"][word="in"]➔ *take in*
- 只有 **word** 属性时可**省略中括号[]**,直接用双引号检索

  "depend" "on" ➔ *depend on*
- **空的[]**可以指示任意字符

  "a" []"of" ➔ *a lot of, a series of, …*
- 将不同属性通过运算符(**和"&"、 或"|"、 否"!="**) 进行组合表达

[word="look"&pos!="N.*"]表示检索的单词为 look,词性不是名词, 即动词 look

- **<s>标示句首,</s>标示句尾**

  <s>"DNA" ➔ "*DNA*"位于句首的句子
- **within s** 放在**检索式末尾**表示出现在**同一句子**中

"set"[]*"up" within s➔ 检索出现在同一句子中的 *set…up*, 其中*为通配符,表示[](即任意字符)出现 0 次或多次

- 检索表达式中常用的**通配符**

| | | |
|---|---|---|
| . | 匹配任意字符或非字符 | "2.4"➔*2a4, 214, 2 4, 2.4, 2-4* |
| * | 重复 0 次或多次 | "*able"➔*able, table, capable, …* |
| .* | 某个字符重复 0 次或更多次 | R.*表示 R 后面有 0 个或多个字符,包括副词 RR, RG, RGQ 等,所以 R.*表示所有副词 |
| + | 重复 1 次或多次 | "A+ "➔ *A, AA, AAA, …* |
| ? | 重复 0 次或 1 次 | "A? "➔0 个或 1 个 A |

| | | |
|---|---|---|
| “ ” | 检索对象必须放” ”里 | [word=”increase”&pos=”N.*”] |
| [] | 分隔符 | 只有 word 属性时可省略[],如”increase”；有词性赋码时不可省 |
| [] | 字符集,用在” ”内 | " [abc] " ➔*a, b, c*　" [1-3] "➔*1, 2, 3* |
| () | 组合，括号中的部分可当作一个符号处理 | "act(ing\|or) "➔*actor, acting* ([pos=”R.*”][pos=”J.*”]) 副词和形容词作为一个整体检索 |
| {} | 表示选择范围 | [pos=”j.*”]{0,2}表示 0-2 个形容词 {0,}表示 0 个或多个 |
| & | 和,并且 | [word=”increase”&pos=”N”] 表示名词词性的 increase |
| \| | 或者 | "analyze\|analyse"➔*analyze, analyse* |
| ! | 不等于,用在等号前面 | [word!=”,”]不是逗号的词 [pos!=”J.*\|N.*\|I.*”] 词性不是形容词、名词、介词的词 |
| ^ | 逻辑非,用在” ”内 | ".*[^t]able" ➔ *able, capable,…* 不含 table |
| <s> | 段首标记 | <s>”However” 表示以 However 开头的句子 |
| </s> | 段尾标记 | [pos=”N.*”]</s> 以名词结尾的句子 |

[word="DNA.*"] ➔ *DNA, DNA-dependent, DNAse…*

[word="DNA.+"]➔*DNA-dependent, DNAse…*不包括 *DNA*

[word="DNA[^-].+"] ➔ *DNAse, DNASTAR…*不包括 *DNA, DNA-dependent*

## 限定条件检索

进入**[Restricted query]**页面,并勾选**子库**,即可得到在指定子库的检索结果

## 拓展功能

**频数解析功能(Frequency breakdown)**：在检索多个目标词或短语时,对目标词项分别计算频数和百分比

- 在**检索结果页面**点击**[New query]**下拉菜单中,选中**[Frequency breakdown]**选项,点击右边的**[Go]**,结果会按搜索到的结果分别报告它们在语料库中出现的频数和百分比。
- 例如：输入"increase|increases"　"in|of",使用频数解析,得到下面的结果：

| No. | Search result | No. of occurrences | Percent |
|---|---|---|---|
| 1 | increase in | 1464 | 71.38% |
| 2 | increases in | 441 | 21.5% |
| 3 | increase of | 131 | 6.39% |
| 4 | increases of | 15 | 0.73% |

比较发现英语母语者 increase in 的搭配用法为 72.3+21.5=93.8%, increase of 的搭配用法仅为 7%

**分布显示功能(Distribution)**：按不同学科领域分别呈现查询结果

- 在**检索结果页面**点击**[New query]**下拉菜单中,选中**[Distribution]**选项,可查看该关键词在**不同学科领域的分布情况**

检索 "limitation|limitations|weakness|weaknesses",使用分布显示,结果如下：

**Based on classification: discipline**

| Category | Words in category | Hits in category | Dispersion (no. files with 1+ hits) | Frequency [↓] per million words in category |
|---|---|---|---|---|
| DWKX | 688000 | 42 | 30 out of 119 | 61.05 |
| NYGC | 483810 | 58 | 32 out of 84 | 119.88 |
| YYLX | 718516 | 61 | 32 out of 102 | 84.9 |
| ZWKX | 786489 | 29 | 18 out of 109 | 36.87 |
| ZYHJ | 537203 | 95 | 33 out of 83 | 176.84 |
| Total: | 5537109 | 453 | 235 out of 838 | 81.81 |

分布数据表明植物科学(ZWKX)最少提及局限和弱点,而资源环境(ZYHJ)科学更多提及

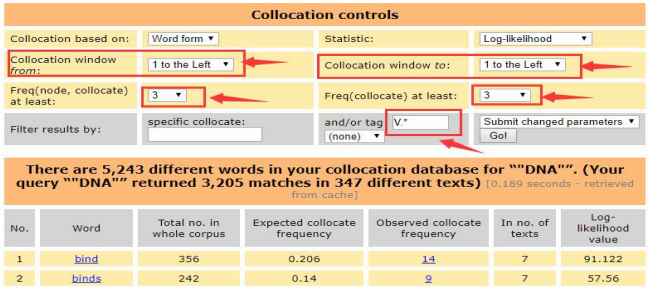
- 学科简写见附表，如 SMKX 表示生命科学
- 若在**[Show as]**一栏中选择 **[Bar Chart]**,可以以柱状图的形式展示结果
- 在章节**子语料库**中检索,可以查看检索词在各个章节部分的出现频率

**搭配功能(Collocations )**：统计特定词语在语料库中的典型搭配。

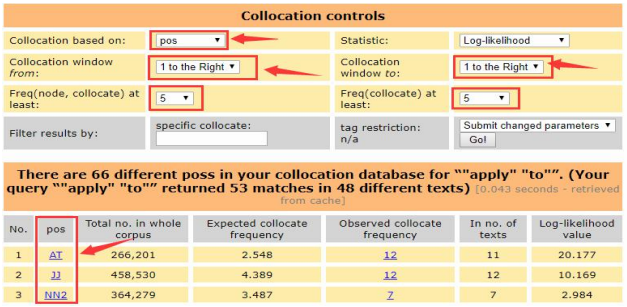
- 在**检索结果页面**点击**[New query]**下拉菜单中,选中**[Collocations]**选项,在下一个界面点击**[Create a collocation database]**,设置搭配检索的条件

| | |
|---|---|
| window from/ to | 指检索目标距离关键词的距离 |
| 1 to the left | 表示关键词左边紧接的单词 |
| Freq | 表示出现的频次条件 |
| and/or tag | 设置词性条件(与其搭配的名词、动词,介词等) |

例：检索”DNA”,并进入搭配功能页面,将搭配检索的条件按照下图所示设置,则表示查询**紧接在 DNA 之前**(1 to the left)、出现**频次不少于 3** 的所有**动词**搭配




例：检索”apply” “to”,搭配检索的条件按照下图所示设置,则表示查询**紧接在 apply to 之后**(1 to the Right)、出现**频次不少于 5** 的搭配词的**词性(pos)**

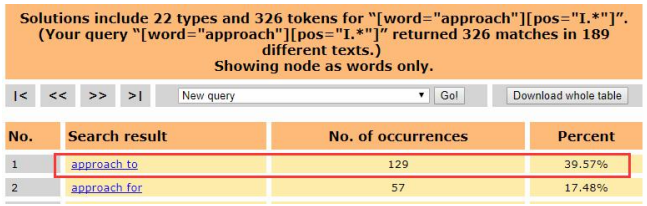



结果显示,apply to 后面与定冠词、形容词、名词搭配,暗示了短语 apply to 中的 to 是介词

## 检索示例

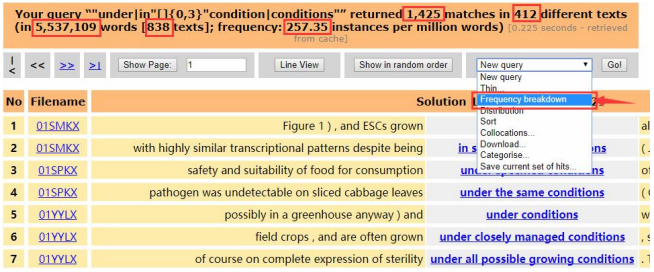
- 例 1：名词 approach 在表示“……的方法”时，与之搭配的介词是什么呢？

检索表达式为：[word="approach"][pos="I.*"],在词性赋码集 CLAWS7 中(见附录),介词的赋码包括 IF、II、IO、IW,所以 I.*表示检索所有形式的介词，即检索单词 approach 与其后的介词。 点击右上角下拉菜单选择 Frequency breakdown，点击 Go，观察结果中 approach 右边第一位搭配频率最高的介词(结果为与 to 搭配 129 词)。




- 例 2：表示“在…条件下”时,与 condition 搭配的介词是 under 居多还是 in 居多？ condition 是单数形式居多还是复数形式居多？

检索表达式: "under|in"[]{0,3}"condition|conditions",其中|表示或者,[]表示任意一个单词,{0,3}表示 condition 前面有 0 到 3 个修饰词,"condition|conditions"表示单数或复数的 condition。

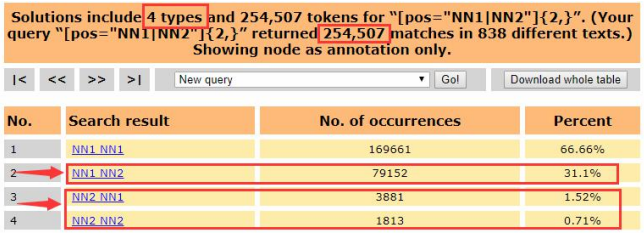



检索结果表明,在 838 个文本(553 万词)中找到 1425 例检索对象,平均每百万词出现 257.35 次。在下拉条中选择 Frequency breakdown,点击 Go,就可以得到检索对象出现的频次和百分比等数据。

- 例 3：“基因网络”的英语表达是 gene networks 还是 genes networks？

检索表达式为: "gene|genes" "networks",结果得到 gene networks 有 14 例,genes networks 有 0 例,表明修饰性的名词 gene 没有复数形式,只有中心名词 network 可以用复数。由此推论,由多个名词组成的名词词组,只有中心名词可以用复数,修饰性的名词没有复数形式。

用表达式[pos="NN1|NN2"]{2,}检索多个名词(NN1 单数或 NN2 复数)组成的词组,其中{2,}表示出现 2 次或以上，在结果页面选择 Frequency breakdown，进入频数解析页面再选择 Frequency breakdown of annotation only，表示只按照检索结果的词性进行频数解析，得到以下结果：

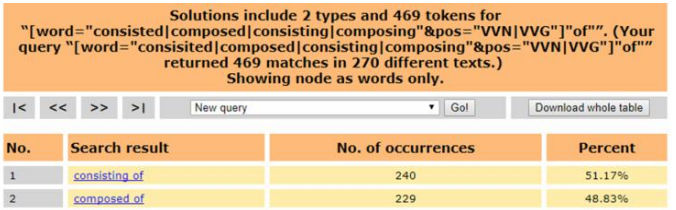



在检索到的 25 万例结果中，只有两个名词组成的词组出现，单数+复数名词的搭配占 31.1%，而复数+单数/复数名词占 1.51+0.71=2.22%(其中大部分为词性注释错误)，由此看出，修饰性的名词极少使用复数形式。

点击词性赋码可以查看对应的单词，如点击 NN1 NN2，可以查看单数+复数名词的搭配，如 protein interactions, target cells, expression profiles 等。

- 例 4：consist 和 compose 均指”由…组成”，两者作后置定语时用主动语态还是被动语态？

检索表达式为：[word="consisted|composed|consisting|composing"&pos="VVN|VVG"]"of"，其中用 pos=”VVN|VVG”限制后置定语的词性为 VVN 过去分词或 VVG 现在分词，检索结果如下：




结果表明，consist 作后置定语的用法为 consisting of，为主动语态，而 compose 作后置定语的用法为 composed of，为被动语态，两者的使用频率相差不大。

## 中英文对照表

| | |
|---|---|
| **(word) token** | 形符/词次 |
| **(word) type** | 类符/词种 |
| **Collocation** | 词语搭配 |
| **Corpus documentation** | 语料库说明文档 |
| **Corpus metadata** | 语料库元信息 |
| **CQP syntax** | 复杂检索语法 |
| **Distribution** | 分布(按语体等分类条件分别呈现结果) |
| **Frequency** | 频数、频率 |
| **Frequency breakdown** | 频数分解、分解频数 |
| **Frequency list** | 词频表、词表 |
| **Keywords** | 主题词 |
| **Log likelihood (ratio)** | 对数似然率(典型词语搭配的统计方法) |
| **Maximum window span** | 最大跨距 (计算搭配时中心词和左右语境词之间的距离) |
| **No. of occurrences** | 出现次数 |
| **Node** | 检索词、中心词、节点词 |
| **Number of hits per page** | 查询结果每页显示的行数 |
| **Query** | 查询、检索 |
| **Restricted query** | 限定条件查询 |
| **Show hits sorted by node** | 查询结果按中心词排序 |
| **Simple query (ignore case)** | 简单查询(不区分大小写) |
| **Word lookup** | 词语相关查询 |

## 附表：子库名称中英文对照表

| 学科子库 | | 章节子库 | |
|---|---|---|---|
| ZWKX | 植物科学 | ABS | 摘要 |
| DWKX | 动物科学 | INT | 引言 |
| SMKX | 生命科学 | MET | 方法 |
| YYLX | 园艺林学 | RES | 结果 |
| NYJJ | 农业经济 | DIS | 讨论 |
| NYGC | 农业工程 | CON | 结论 |
| SCKX | 水产科学 | | |
| SPKX | 食品科学 | | |
| ZYHJ | 资源环境 | | |

## 附录：CLAWS 7　赋码集

| TAG | Example |
|---|---|
| APPGE | possessive pronoun, pre-nominal (e.g. my, your, our) 名词前的所有格 |
| AT | article (e.g. the, no) 定冠词 |
| AT1 | singular article (e.g. a, an, every) 不定冠词 |
| BCL | before-clause marker (e.g. in order (that),in order (to)) 从句前的标志 |
| CC | coordinating conjunction (e.g. and, or)并列连词 |
| CCB | adversative coordinating conjunction ( but)转折连词 |
| CS | subordinating conjunction (e.g. if, because, unless, so, for) 从属连词 |
| CSA | as (as conjunction) 连词 as |
| CSN | than (as conjunction) 连词 than |
| CST | that (as conjunction) 连词 that |
| CSW | whether (as conjunction) 连词 wether |
| DA | after-determiner or post-determiner capable of pronominal function (e.g.such, former, same) 起名词作用的后限定词 |
| DA1 | singular after-determiner (e.g. little, much)后限定词单数形式 |
| DA2 | plural after-determiner (e.g. few, several, many) 后限定词复数形式 |
| DAR | comparative after-determiner (e.g. more, less, fewer) 后限定词比较形式 |
| DAT | superlative after-determiner (e.g. most, least, fewest) 后限定词最高形式 |
| DB | before determiner or pre-determiner capable of pronominal function (all, half) 起名词作用的前限定词 |
| DB2 | plural before-determiner ( both)前限定词复数形式 |
| DD | determiner (capable of pronominal function) (e.g any, some)限定词 |
| DD1 | singular determiner (e.g. this, that, another)单数限定词 |
| DD2 | plural determiner ( these,those)复数限定词 |
| DDQ | wh-determiner (which, what) WH 限定词 |
| DDQGE | wh-determiner, genitive (whose) WH 限定词所有格 |
| DDQV | wh-ever determiner, (whichever, whatever) WH 强调限定词 |
| EX | existential there 存在句 there |
| FO | formula 公式 |
| FU | unclassified word 未分类词 |
| FW | foreign word 外来词汇 |
| GE | germanic genitive marker - (' or's) 所有格标记 |
| IF | for (as preposition) 介词 for |
| II | general preposition 普通介词 |
| IO | of (as preposition) 介词 of |
| IW | 介词 with, without (as prepositions) |
| JJ | general adjective 普通形容词 |
| JJR | general comparative adjective (e.g. older, better, stronger) 形容词一般比较级 |
| JJT | general superlative adjective (e.g. oldest, best, strongest) 形容词一般最高级 |
| JK | catenative adjective (able in be able to, willing in be willing to) 形容词短语 |
| MC | cardinal number,neutral for number (two, three..)基数词 |
| MC1 | singular cardinal number (one)单数基数词 |
| MC2 | plural cardinal number (e.g. sixes, sevens)复数基数词 |
| MCGE | genitive cardinal number, neutral for number (two's, 100's)基数词所有格 |
| MCMC | hyphenated number (40-50, 1770-1827)连字符连接的数字 |
| MD | ordinal number (e.g. first, second, next, last)序数词 |
| MF | fraction,neutral for number (e.g. quarters, two-thirds)分数 |
| ND1 | singular noun of direction (e.g. north, southeast) 方向名词单数 |
| NN | common noun, neutral for number (e.g. sheep, cod, headquarters) 普通名词 |
| NN1 | singular common noun (e.g. book, girl)单数普通名词 |
| NN2 | plural common noun (e.g. books, girls)复数普通名词 |
| NNA | following noun of title (e.g. M.A.)称呼之后的名词 |
| NNB | preceding noun of title (e.g. Mr., Prof.)称呼之前的名词 |
| NNL1 | singular locative noun (e.g. Island, Street)单数方位名词 |
| NNL2 | plural locative noun (e.g. Islands, Streets)复数方位名词 |
| NNO | numeral noun, neutral for number (e.g. dozen, hundred)数量词 |
| NNO2 | numeral noun, plural (e.g. hundreds, thousands)复数数量词 |
| NNT1 | temporal noun, singular (e.g. day, week, year)单数时间名词 |
| NNT2 | temporal noun, plural (e.g. days, weeks, years)复数时间名词 |
| NNU | unit of measurement, neutral for number (e.g. in, cc)度量单位词 |
| NNU1 | singular unit of measurement (e.g. inch, centimetre)单数度量单位词 |
| NNU2 | plural unit of measurement (e.g. ins., feet)复数度量单位词 |
| NP | proper noun, neutral for number (e.g. IBM, Andes)专有名词 |
| NP1 | singular proper noun (e.g. London, Jane, Frederick)单数专有名词 |
| NP2 | plural proper noun (e.g. Browns, Reagans, Koreas)复数专有名词 |

| | |
|---|---|
| **NPD1** | singular weekday noun (e.g. Sunday)单数星期名词 |
| **NPD2** | plural weekday noun (e.g. Sundays)复数星期名词 |
| **NPM1** | singular month noun (e.g. October)单数月份名词 |
| **NPM2** | plural month noun (e.g. Octobers)复数月份名词 |
| **PN** | indefinite pronoun, neutral for number (none)不定代词 |
| **PN1** | indefinite pronoun, singular (e.g. anyone, everything, nobody, one)单数不定代词 |
| **PNQO** | objective wh-pronoun (whom)宾格 WH 代词 |
| **PNQS** | subjective wh-pronoun (who)主格 WH 代词 |
| **PNQV** | wh-ever pronoun (whoever) WH 代词强调形式 |
| **PNX1** | reflexive indefinite pronoun (oneself)反身不定代词 |
| **PPGE** | nominal possessive personal pronoun (e.g. mine, yours) 名词所有格人称代词 |
| **PPH1** | 3rd person sing. neuter personal pronoun (it)第三人称单数中性人称代词 |
| **PPHO1** | 3rd person sing. objective personal pronoun (him, her)第三人称单数宾格人称代词 |
| **PPHO2** | 3rd person plural objective personal pronoun (them) 第三人称复数宾格人称代词 |
| **PPHS1** | 3rd person sing. subjective personal pronoun (he, she) 第三人称单数主格人称代词 |
| **PPHS2** | 3rd person plural subjective personal pronoun (they) 第三人称复数主格人称代词 |
| **PPIO1** | 1st person sing. objective personal pronoun (me) 第一人称单数宾格人称代词 |
| **PPIO2** | 1st person plural objective personal pronoun (us) 第一人称复数宾格人称代词 |
| **PPIS1** | 1st person sing. subjective personal pronoun (I) 第一人称单数主格人称代词 |
| **PPIS2** | 1st person plural subjective personal pronoun (we) 第一人称复数主格人称代词 |
| **PPX1** | singular reflexive personal pronoun (e.g. yourself, itself) 单数反身人称代词 |
| **PPX2** | plural reflexive personal pronoun (e.g. yourselves, themselves) 复数反身人称代词 |
| **PPY** | 2nd person personal pronoun (you)第二人称代词 |
| **RA** | adverb, after nominal head (e.g. else, galore)名词中心词后的副词 |
| **REX** | adverb introducing appositional constructions (namely, e.g.)引导同位语结构的副词 |
| **RG** | degree adverb (very, so, too)程度副词 |
| **RGQ** | wh- degree adverb (how)疑问程度副词 |
| **RGQV** | wh-ever degree adverb (however)强调程度副词 |
| **RGR** | comparative degree adverb (more, less)程度副词比较级 |
| **RGT** | superlative degree adverb (most, least)程度副词最高级 |
| **RL** | locative adverb (e.g. alongside, forward)地点副词 |
| **RP** | prep. adverb, particle (e.g about, in) 介词、副词、小品词 |
| **RPK** | prep. adv., catenative (about in be about to) 短语中介、副词 |
| **RR** | general adverb 普通副词 |
| **RRQ** | wh- general adverb (where, when, why, how) WH 普通副词 |
| **RRQV** | wh-ever general adverb (wherever, whenever) WH-EVER 普通副词 |
| **RRR** | comparative general adverb (e.g. better, longer)普通副词比较级 |
| **RRT** | superlative general adverb (e.g. best, longest)普通副词最高级 |
| **RT** | quasi-nominal adverb of time (e.g. now, tomorrow)准名词性时间副词 |
| **TO** | infinitive marker (to)不定式标记 |
| **UH** | interjection (e.g. oh, yes, um)感叹词 |
| **VB0** | be, base form (finite i.e. imperative, subjunctive)动词原型 BE |
| **VBDR** | were 系动词过去式 were |
| **VBDZ** | was 系动词过去式 was |
| **VBG** | being 现在分词 being |
| **VBI** | be, infinitive (To be or not... It will be ..)动词不定式 |
| **VBM** | am 系动词 am |
| **VBN** | been 过去分词 been |
| **VBR** | are 系动词 are |
| **VBZ** | is 系动词 is |
| **VD0** | do, base form (finite)动词原形 DO |
| **VDD** | did 动词过去式 did |
| **VDG** | doing 现在分词 doing |
| **VDI** | do, infinitive (I may do... To do...)不定式 do |
| **VDN** | done 过去分词 done |
| **VDZ** | does 动词 does |
| **VH0** | have, base form (finite) 动词原形 HAVE |
| **VHD** | had (past tense)过去式 had |
| **VHG** | having 现在分词 having |
| **VHI** | have, infinitive 不定式 have |
| **VHN** | had (past participle)过去分词 had |
| **VHZ** | has 动词 has |
| **VM** | modal auxiliary (can, will, would, etc.)情态助动词 |
| **VMK** | modal catenative (ought, used)情态动词 |
| **VV0** | base form of lexical verb (e.g. give, work)实义动词原形 |
| **VVD** | past tense of lexical verb (e.g. gave, worked)实义动词过去式 |
| **VVG** | -ing participle of lexical verb (e.g. giving, working)实义动词 ing 分词形式 |
| **VVGK** | -ing participle catenative (going in be going to)短语中的 ing 分词 |
| **VVI** | infinitive (e.g. to give... It will work...)不定式 |
| **VVN** | past participle of lexical verb (e.g. given, worked)实义动词过去分词 |
| **VVNK** | past participle catenative (e.g. bound in be bound to)短语中的过去分词 |
| **VVZ** | -s form of lexical verb (e.g. gives, works)实义动词-s 形式 |
| **XX** | 否定词 not, n't |
| **ZZ1** | singular letter of the alphabet (e.g. A,b)单数字母 |
| **ZZ2** | plural letter of the alphabet (e.g. A's, b's)复数字母 |

2019.12.25 整理