

CCA User Guide

CCA offers free online collocation search service to teachers, students and researchers who are interested in appropriate use of modern Chinese collocations. Currently our database contains over 300,000 Modern Chinese collocations automatically extracted from CTC, a text corpus for second language acquisition built by Beijing Normal University, and over 1,450,000 collocations from Chinese Wikipedia Corpus. We define nine types of collocations with 26 forms in the following table.

Type	Form	Example
Classifier-Noun	classifier+noun	个人 (ge person) – person
Preposition-Postposition	preposition+X+postposition	在 X 上 (zai X shang) – on X
Preposition-Verb	preposition+X_verb	把 X 解决 (ba X solve) – solve X
	preposition+X+verb+particle	把 X 吃了 (ba X eat le) – eat something
	preposition+X+verb+particle1+particle2	把 X 吃完了 (ba X eat up le) – have eaten up X
Predicate-Complement	predicate+(X)+complement	吃饱 (eat full) – eat to be full
	predicate+complement+particle	快乐多了 (happy more le) – much happier
	predicate+adverb+complement	看不出来 (recognize not out) – can not recognize
	predicate+particle+(X)+complement	玩得愉快 (play de happy) – play happily
	predicate+particle+adverb+complement	玩得非常愉快 (play de quite happy) – play quite happily
	predicate+numeral+complement	试三遍 (try three times) – try three times
	predicate+particle+numeral+complement	试了三遍 (try le three times) – have tried three times
	predicate+adjective+complement	穿漂亮一点 (dress beautiful a bit) – dress a bit more beautifully
	predicate+particle+adjective+ complement	穿得漂亮一点 (dress de beautiful a bit) – dress a bit more beautifully
Verb-Object	verb+object	喜欢看书 (like reading) – like reading
	verb+particle+object	唱着歌 (sing zhe song) – singing a song
	verb+particle1+particle2+object	学到了技术 (learn dao le technology) – have learnt the technology
Subject-Predicate	subject+(X)+predicate	歌曲流行 (song popular) – the song is popular
	subject+(X)+predicate+particle	我经历过 (I experience guo) – I have experienced
	subject+(X)+predicate+particle1+ particle2	戒指找回来了 (ring find back le) – the ring is found
Adjective-Noun	adjective+(X)+noun	著名大学 (famous university) – famous university
	adjective+(X)+de+(X)+noun	卑鄙的人 (mean de people) – mean people
Adverb-Predicate	adverb+(X)+verb	突然改变 (suddenly change) – change suddenly
	adverb+de+verb	有效地提高 (effective de improve) – improve effectively
	adverb+(X)+adjective	非常漂亮 (very beautiful) – very beautiful
Connectives	conjunction/adverb+X+conjunction/adverb	虽然 X 但是 (although X but) – although

The collocation data in CCA is automatically extracted from the text corpus and may inevitably contain mistakes. We highly appreciate users' questions and suggestions regarding the data, the UI/UX design and anything that could help us to improve CCA. Please contact us via the following email address if you find collocation errors or have questions/advices:

Email: irishu@mail.bnu.edu.cn (Dr. Renfen Hu)

Additionally, we provide access to the raw data of our Chinese Collocation Knowledge Base. For more information, please refer to the collocation data documentation on GitHub¹.

We sincerely hope to offer convenient and reliable collocation search service to assist the Chinese vocabulary learning.

Thank you!

CCA Team

July, 2023

¹ https://github.com/iris2hu/Chinese-collocation-complexity/blob/main/collocation_data/collocation_data.md

CCA 使用说明

中文搭配助手(CCA)应用语言信息处理技术从语料库中自动抽取搭配信息，向用户提供现代汉语搭配在线检索服务，以辅助汉语教学及相关领域研究。为了适应二语教学的领域需求，搭配数据抽取主要基于北京师范大学构建的汉语国际教育动态语料库（收录经典汉语教材及新 HSK 真题文本语料）。目前，CCA 数据库包含 30 余万条取自汉语国际教育动态语料库的中文搭配及其频次、互信息、上下文信息等属性。此外，我们还从中文维基百科语料库中抽取了约 145 万条中文搭配数据，作为更为全面的搭配信息参考。

考虑到汉语的句法语义特性，尤其是汉语中特有虚词的搭配用法，我们定义了九种搭配类型，分别为：量名搭配、动宾搭配、主谓搭配、定中搭配（特指形容词与名词搭配）、状中搭配（特指副词与动词或形容词搭配）、介谓搭配、述补搭配、介方搭配、关联词（连词、副词）搭配。用户在检索时可指定一种或多种类型，九种搭配的形式及示例可参见下表：

类型	形式	示例
量名搭配	量词+名词	个人；头牛
动宾搭配	动词+宾语	喜欢 看书
	动词+助词+宾语	唱着 歌
	动词+助词 1+助词 2+宾语	学 到了 技术
主谓搭配	主语+谓词	歌曲 流行
	主语+谓词+助词	我 经历过
	主语+谓词+助词 1+ 助词 2	戒指 找 回来了
定中搭配	形容词+名词	著名 大学
	形容词+的+名词	卑鄙 的人
状中搭配	副词+动词	突然 改变
	副词+地+动词	有效 地 提高
	副词+形容词	非常 漂亮
介谓搭配	介词+X+谓词	把 X 解决
	介词+X+谓词+助词	把 X 吃了
	介词+X+谓词+助词 1+助词 2	把 X 吃完了
述补搭配	谓词+补语	吃 饱
	谓词+补语+助词	快乐 多了
	谓词+副词+补语	看 不出来
	谓词+助词+补语	玩 得 愉快
	谓词+助词+副词+补语	玩 得 非常 愉快
	谓词+数词+量词	试 三 遍
	谓词+助词+数词+量词	试 了 三 遍
	谓词+形容词+补语	穿 漂亮 一点
谓词+助词+形容词+补语	穿 得 漂亮 一点	
介方搭配	介词+X+方位词	在 X 上；从 X 起
关联词搭配	连词/副词+X+连词/副词	不但 X 而且；先 X 再

需要指出的是，当前版本的搭配数据均由相应算法从语料库中自动抽取得到，未经人工校对，因而存在错误与纰漏。此外，系统处于试运行阶段，尚有诸多不完善之处。用户在试用期间如发现搭配数据中的错误，或有任何形式的问题或建议，都欢迎通过以下联系方式向我们提出：

邮箱：irishu@mail.bnu.edu.cn（胡韧奋博士）

此外，为服务于科研用途，我们也提供了中文搭配知识库的原始数据下载，具体可参考 **Github** 项目的数据说明²。

我们将致力于持续提升搭配数据的规模与质量，改进完善 CCA 的功能，向用户提供准确、流畅的中文搭配检索服务。

CCA Team
2023 年 7 月

² https://github.com/iris2hu/Chinese-collocation-complexity/blob/main/collocation_data/collocation_data.md