

# 语料库与中文教学

## Corpus and Chinese L2 Teaching

主讲人: Chengzhi Chu 储诚志, University of California, Davis  
主持人: Jun Da 笮骏, Middle Tennessee State University



## 纪念吕必松先生 (1935-2017)

作为中国对外汉语教学学科的奠基人，吕必松先生也是将语料库理念和信息处理技术引入汉语作为第二语言教学与研究的重要推手。《现代汉语频率词典》、《汉语中介语语料库系统》等代表性基础项目，以及北京语言学院语言信息处理研究所（今北京语言大学信息科学学院前身）的建立，都是因他的远见和规划而得以完成。

“卻顧所來徑，蒼蒼橫翠微”



核心问题：

语料库有何功能及用途？

中文教学中如何利用语料库？



# 语料库：自然语言实例的电子化资源库

自然语言：真实语料，不是研究者编的，可能是驳杂的，质量不均衡

——篇章(文字书写的)

——话语(语音文档，目前一般也有相应的文字转写，以便处理)

实例：定向或随机搜集，有计划或无计划。样本：

——活语料无尽，只能是经过选择(通常是抽样)的

——死语料可以穷尽(如“先秦典籍”)

资源：是为其他目的服务的，建库本身不是目的；

**backward design**: 根据目的收集资源

库：通常都是大规模的；

有内部的数据处理规范和库结构，便于数据分析和提取；

——标注语篇属性

——熟语料：进行字、词、句、段不同等级的加工和标记

检索系统：面向用户

数据更新与扩充系统：面向研究者



## 语料库的主要分类：

母语语料库 vs. 中介语/学习者语料库

单语 vs. 平行语料库 (翻译)

通用平衡语料库 vs. 专门语料库

(口语、书面语；报刊、小说、法律文件；先秦汉语、唐诗、佛经；中文教材…)

生语料库 vs. 熟语料库 (parsed corpus, tree bank)



# 语料库在二语教学中的用处

基础工程：制定教学大纲、编写教材、教学辅助材料和词典等

日常教学：随时都有用，常常应该用。见下面举例

学生学习：很多时候好的语料库比好的词典对学生更有帮助，尤其是中高级阶段以后

实际问题举例：

1. 现代汉语的常用词、常用字有多少？是哪些？
2. 中文一年级应该教多少字、多少词？教哪些？
3. 《中文听说读写》的字、词、语法点的选用和对话、短文的编写做得怎么样？跟《新实用汉语课本》相比呢？学了IC-1可不可以改用 Chinese Link 2？
4. 《汉语风》500词级的读物适合你们班的学生阅读吗？
5. 学生上了一年中文课，学过了什么？学到了什么？能说什么水平的话？对“把”字句掌握到了什么程度？
6. “能”和“能够”有什么不同？“提高”和“提升”有什么不同？“promote”和“提升”有什么不同？“睡觉”和“唱歌”都是离合词吗？
7. “这件事我同意老王”能不能说？“我有去上海”的说法到底流行开了没有？谁接受谁不接受？“把”字句中“把”的宾语真的只能是有定的吗？
8. 孔子曰过某句话吗？
9. 《七步诗》怎么翻译？



## 使用语料库的两个研究路径

Corpus-based Approach (基于语料库的研究) :

用语料来验证 (证实或证伪) 某种语言直觉或理论假设。

如: “把” 带不带无定宾语 (储诚志 2021) ;

“女士、女子、女人、妇女” 的语体差异 (张正生 2021) 。

Corpus-driven Approach (语料库驱动的研究) :

通过对语料库文本不同层次的观察分析来发现规律。

如: “常常” 和 “往往” 的不同;

IC的字、词、句、篇分析;

学生对“把”字句的习得情况。

这两个路子通常是互相交织的。



# 语料库在研究和教学中的应用举例





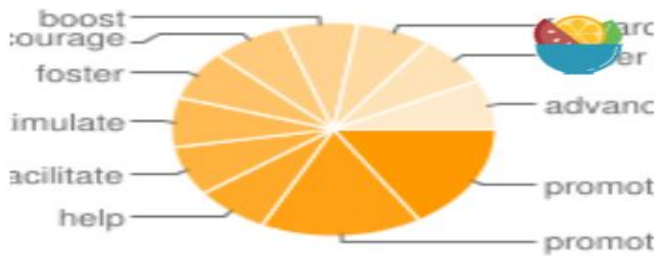
# “促进”：什么意思？怎么使用？

## 查词典不如查语料库

### 促进：

n. promotion  
 v. promote help facilitate  
stimulate foster encourage  
boost further

### 释义分布



### 后面的常见名词

和平 协会 委员会 作用  
 方面 经济 会议 关系 贸  
 易 事业

### 前面的常见副词

进一步 相互 大大 共同  
 努力 大力 这样 反过来  
 必将 必定

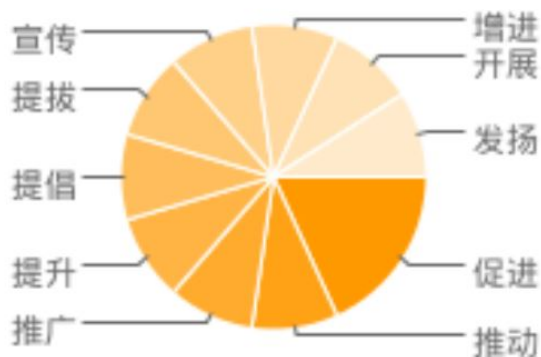
1. We haven't found a better tool for initial sketches than a simple whiteboard. Working at a whiteboard **promotes** collaboration and discussion and, of course, everything is easy to erase and redraw.  
 一块简单的写字板是草拟框架再好不过的工具了，在写字板上的工作可以**促进**协作和讨论。还有，上面的内容很容易被擦掉或者重画。  
 -- 来源 -- About Face 3交互设计精髓 [好评\(13\)](#) [差评\(2\)](#)
2. To state the obvious, then, to make people more productive and happy, it behooves us to design interactive products to promote and enhance flow, and for us to go to great pains to avoid any potentially flow-disturbing behavior.  
 这显然也会让人们更加高效并享受其中，因此我们在设计交互产品时要**促进**和增强流，即使有极大的困难也要避免任何可能打断流的行为。  
 -- 来源 -- About Face 3交互设计精髓 [好评\(19\)](#) [差评\(9\)](#)
3. After you have a good list of top-level functional and data elements, you can begin to group them into functional units and determine their hierarchy. <sup>3</sup> Because these elements facilitate specific tasks, the idea is to group elements to best **facilitate** the persona's flow (see Chapter 10) both within a task and between related tasks.  
 当你有了完善的高层次的功能和数据元素之后，便可以开始将其分组，并决定其层次<sup>3</sup>。因为这些元素都将有助于具体的任务，因此元素分组的目的在于更好地在任务中和任务间来帮助（**促进**）人物角色的操作流程（参见第10章）。  
 -- 来源 -- About Face 3交互设计精髓 [好评\(13\)](#) [差评\(6\)](#)
4. Their frequency of use both **justifies** and requires the memorization.  
 频繁的使用需要记忆，也**促进**了记忆。  
 -- 来源 -- About Face 3交互设计精髓 [好评\(11\)](#) [差评\(4\)](#)
5. If the goal **is** consistently **positive** customer interactions, the verbal, visual, and behavioral brand messages must be consistent.  
 如果目标始终都是**促进**良好积极的顾客交互，那么语言、视觉和行为的品牌信息就必须一致。  
 -- 来源 -- About Face 3交互设计精髓 [好评\(12\)](#) [差评\(6\)](#)
6. **The manipulation** affordances of contemporary faux three-dimensional buttons have **prompted** their widespread use. It's a good thing—so why not use it a lot?  
 当代人造三维按钮的操作启示**促进**了它的广泛使用，它是好东西，那么为什么不多用它？  
 -- 来源 -- About Face 3交互设计精髓 [好评\(15\)](#) [差评\(9\)](#)
7. The value of qualitative studies is not limited to **helping support** the design process. In our experience, spending the time to understand the user population as human beings can provide valuable business insights that are not revealed through traditional market research.  
 定性研究的价值不仅在于**促进**设计流程，根据我们的经验，花时间理解具体用户还能够提供传统市场研究所无法发现的宝贵的商业见解。  
 -- 来源 -- About Face 3交互设计精髓 [好评\(12\)](#) [差评\(7\)](#)
8. We've found it useful to render Framework illustrations in a visual style that suggests the sketchiness of the proposed solutions (recall that rough sketches tend to do a better job promoting discourse about design).  
 在我们自己的经验中，发现勾画出可以表现方案梗概的框架草图非常有帮助（我们先说过，草图可以很好地**促进**我们展开对设计的讨论）。  
 -- 来源 -- About Face 3交互设计精髓 [好评\(12\)](#) [差评\(8\)](#)

# “promote” 中文怎么说？ 查词典不如查语料库

**promote**: [ prə'məʊt ] 

v. 促进 推动 推广 提升 提倡  
提拔 宣传 增进

## 释义分布



jukuu.com/search.php?q=promote



promote

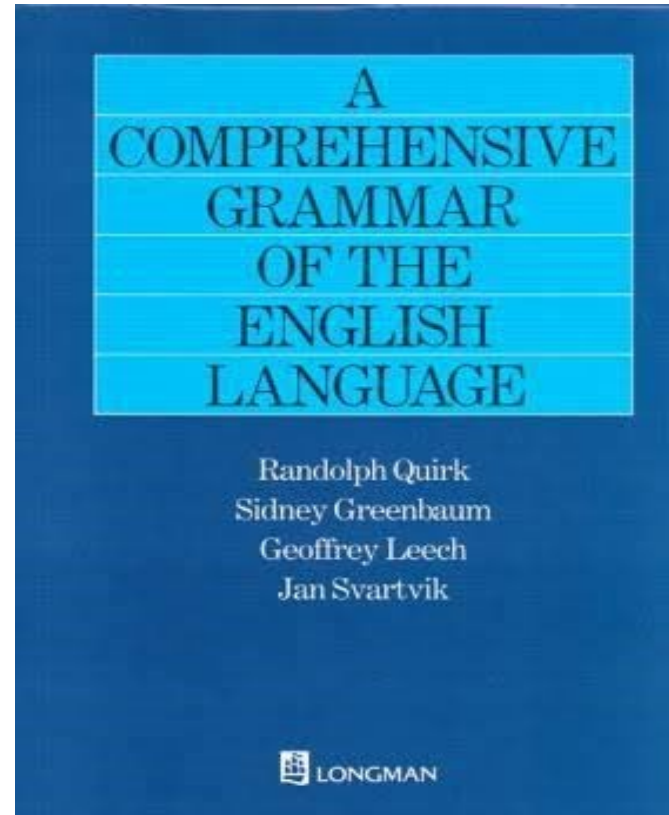
句酷一下

帮助 [单词辨析](#)

1. form an association to **promote** social welfare [详情](#)  
成立一个**促进**社会福利的协会  
-- 来源 -- 英汉 - 辞典例句 [好评\(20\)](#) [差评\(0\)](#)
2. Ease-of-use is obviously important, but it doesn't **promote** our craft to attribute its success to metaphysics.  
易用显然非常重要, 但不能把**将**界面设计的成功**归于**形而上学。  
-- 来源 -- About Face 3交互设计精髓 [好评\(17\)](#) [差评\(0\)](#)
3. **promote** an active campaign  
促进一项积极的运动  
-- 来源 -- 英汉 - 辞典例句 [好评\(16\)](#) [差评\(0\)](#)
4. The policy will **promote** rather than hinder reform. [详情](#)  
这项政策将**促进**而不是妨碍改革。  
-- 来源 -- 英汉 - 辞典例句 [好评\(16\)](#) [差评\(0\)](#)
5. **promote** the cause of cremation  
**促进**火化工作  
-- 来源 -- 英汉 - 辞典例句 [好评\(11\)](#) [差评\(0\)](#)
6. To state the obvious, then, to make people more productive and happy, it behooves us to design interactive products to **promote** and enhance flow, and for us to go to great pains to avoid any potentially flow-disturbing behavior.  
这显然也会让人们更加高效并享受其中, 因此我们在设计交互产品时要**促进**和增强流, 即使有极大的困难也要避免任何可能打断流的行为。  
-- 来源 -- About Face 3交互设计精髓 [好评\(19\)](#) [差评\(9\)](#)
7. And the show will help **promote** it.  
而这次展览会有利于**促销**它。  
-- 来源 -- 英汉 - 辞典例句 [好评\(11\)](#) [差评\(1\)](#)
8. Anhydride-type curing agents require the application of heat to **promote** polymerization.  
[详情](#)  
酞酐类固化剂则要求加热以**促进**聚合反应。  
-- 来源 -- 英汉 - 辞典例句 [好评\(11\)](#) [差评\(1\)](#)
9. Supply-side economics was supposed to **promote** savings, investment, and entrepreneurial creativity. [详情](#)  
供应经济学政策的经济应该能够**提高**存款, 刺激投资和企业创造力。  
-- 来源 -- 英汉 - 辞典例句 [好评\(10\)](#) [差评\(0\)](#)
10. **promote** a bill in Parliament  
努力**促使**法案在国会**通过**  
-- 来源 -- 英汉 - 辞典例句 [好评\(10\)](#) [差评\(1\)](#)



应用举例: R. Quirk et al (1985) English grammar  
based on the Survey of English Usage (SEU) :



应用举例:

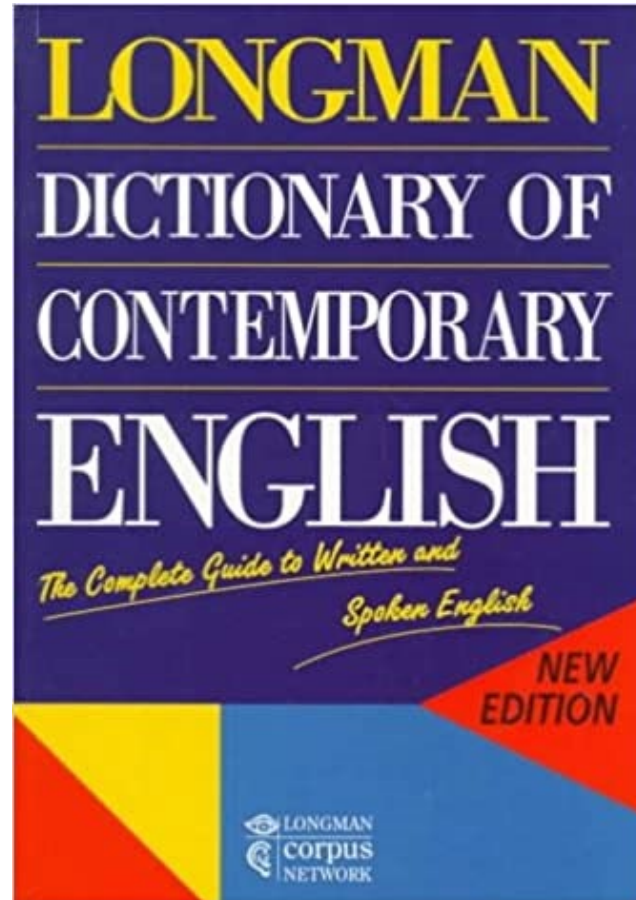
Douglas Biber: Grammar、 discourse and formulaic language and register variation in English

- Biber, D., S. Johansson, G. Leech, S. Conrad, E. Finegan. 1999. *The Longman grammar of spoken and written English*. London: Longman.
- Biber, D. 2006. *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.





应用举例：



TCLT Tech-Talk Series  
@<http://www.tclt.us/talk/>

# 语料库的应用举例 (1)

## 基于“汉语中介语语料库系统” (储诚志、陈小荷等 1995) 的研究

世界汉语教学 1995 年第 4 期(总第 34 期)

### “汉语中介语语料库系统”研制成功

本刊记者

中国国家教育委员会“八五”人文、社会科学科研规划项目,中国国家对外汉语教学领导小组办公室“八五”科研规划项目,北京语言学院“八五”科研规划重点项目——“汉语中介语语料库系统”于 1995 年 11 月 15 日在北京语言学院通过了专家鉴定。

“汉语中介语语料库系统”于 1992 年底立项,1993 年初组建课题组,由北京语言学院教务处组织实施,储诚志为课题组长,陈小荷为副组长,张旺熹、张伟、魏苹、朱旗为课题组成员。课题组还聘请了邓守信、胡明扬、简慕善、柯彼德、黎天睦、李英哲、刘君若、卢绍昌、陆俭明、吕必松、郑锦全等海内外 11 位从事汉语教学与研究的著名学者作为学术顾问。

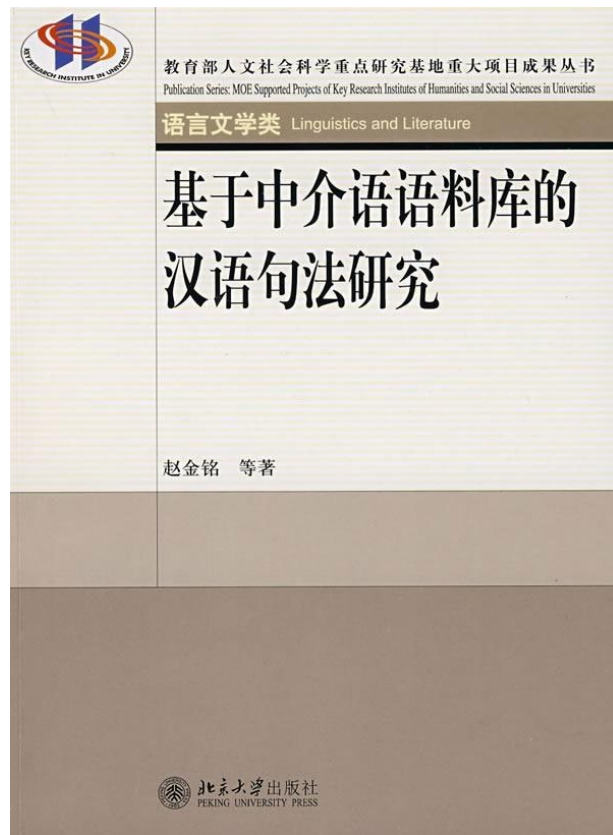
“汉语中介语语料库系统”的研制工作从 1993 年初开始到 1995 年 11 月结束,历时近 3 年时间。其研制工作的基本流程是:总体设计→语料收集→文字预处理→语料录入与校对→语料和作者属性标注→属性登录与校对→文本过滤→文字预处理信息登录→抽样→断句与断句校对→分词与分词校对→词性标注与词性校对→检索系统设计。在研制工作的每一环节,又遵循实验研究→方案制定→全面加工的工作步骤进行操作。

“汉语中介语语料库系统”从北京语言学院、首都师范大学、中国人民大学、北京大学、北京师范大学、安徽师范大学等 9 所高校收集了 96 个国家和地区的 1635 位外国学生成篇成段的汉语作文或练习材料 5774 篇,共计 3528988 字。以这些语料为抽样总体,根据抽样方案进行抽样,最后抽取了 740 位作者写作的 1731 篇共计 1041274 字的语料作为样本加工入库。

“汉语中介语语料库系统”登录了全部语料的 23 个属性:1. 作者姓名,2. 性别,3. 年



# 基于“汉语中介语语料库系统”（储诚志、陈小荷等 1995）的研究



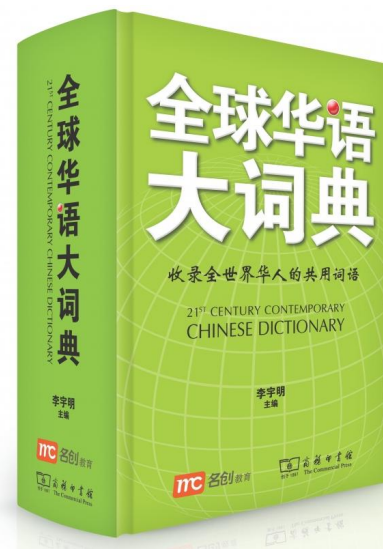


## 基于语料库的研究举例（2）：《全球华语大辞典》的编写

李宇明主编（北美组：储诚志(PI), 王志军, 张英宇）  
北京商务印书馆 & 新加坡名创 2016

一部反映世界主要华语区当代华语词汇面貌的语文辞书。注重实用性，力求促进不同华人社区之间的交流，在华语使用中起协调作用。编写过程中充分吸收现有学术成果，反映当代学术研究的最高水平。充分利用了计算机和互联网等现代化手段，提高了辞书编纂的技术水平。

收录华语通用词语和特有词语约88400条  
注明特有词语的使用地区和异名词语  
同异兼收，展示丰富多样的华语词汇  
知同察异，共建包容和谐的语文生活





## 基于语料库的研究举例（3）： 《汉语风》中文分级系列读物的研制与写作

刘月华、储诚志（2007-2019）  
储诚志、赵绍玲（2019-2021）  
北京大学出版社

《汉语风》(*Chinese Breeze*)是一套大型中文分级泛读系列丛书。这套丛书以“学习者通过轻松、广泛的阅读提高语言的熟练程度、培养语感、增强对中文的兴趣和学习自信心”为基本理念，根据难度分为8个等级，每一级8-10册，共60余册，每册8,000至30,000字。丛书的读者对象为大学和中学里从初级一直到高级（大致掌握300-4,500常用词）水平的中文学生，以及水平在此之间的其他中文学习者。



# Design and Features of 《汉语风》 (*Chinese Breeze*)

## 一、等级完备，方便选择

设计的八个语言等级，能满足不同程度的中文学习者的需要，使他们都能找到适合自己语言水平的读物。

八个等级的读物所使用的基本词汇数目如下：

第1级： 300 基本词	第5级： 1,500 基本词
第2级： 500 基本词	第6级： 2,100 基本词
第3级： 750 基本词	第7级： 3,000 基本词
第4级： 1,100 基本词	第8级： 4,500 基本词



## Design and Features of 《汉语风》 (Chinese Breeze)

### 三、词汇实用，反复重现。

各等级读物所选用的基础词语是该水平等级的学习者的中文交际中最需要、最常用的。

为研制《汉语风》各等级的基础词表，《汉语风》工程首先建立了两个语料库：一个是大规模的当代中文书面语和口语语料库，一个是以世界上不同地区有代表性的40余套中文教材为基础的教材语言库。然后根据不同的交际语域和使用语体对语料样本进行分层标注，再根据语言学习的基本阶段对语料样本分别进行分层统计和综合统计，最后得出符合不同学习阶段需要的不同的词汇有用度表，以此作为《汉语风》等级词表的基础。此外，《汉语风》等级词表还参考了美国、中国大陆、台湾、香港和英国等所建的10余个当代中文语料库的词语统计结果。

以全新的理念和方法研制的《汉语风》分级基础词表，力求既具有较高的交际实用性，也能与学生所用的教材保持高度的相关性。

此外，《汉语风》的各级基础词语在读物中都通过不同的语境反复出现，以促进语言的学习。



## Design and Features of 《汉语风》 (Chinese Breeze)

### 四、易读易懂，生词率低。

《汉语风》严格控制读物的词汇分布、语法难度、情节开展和文化负荷，使读物易读易懂。

在较初级的读物中，生词的密度严格控制在不构成理解障碍的的1.5%到2%之间，而且每个生词（本级基础词语之外的词）在一本读物中初次出现的当页用脚注做出简明注释，并在以后每次出现时都用相同的索引序号进行通篇索引，篇末还附有生词总索引，以方便学生查找，帮助理解。



## 基于语料库的研究举例（4）： “把”字宾语的有定与无定（储诚志 2021）

在介绍“把”字句的时候，**中文教材里一般都特别说明：“把”字句的宾语是“有定的”，有的说必须是有定的，有的说一般是有定的（对于“有定”这个语法术语，教材中常常说成“确指”“特指”或“某确定事物”等）。中文教师通常也是遵循这一认识进行“把”字句的教学。例如：**

“‘把’的宾语...必须是特指的。”

（杨寄洲等 2006,《汉语教程》第2册第49页）

“The object in the ‘把’ sentence is usually definite in the speaker’s mind. Therefore, we cannot say ‘我先把一本书还了。’ We can only say ‘我先把那本书还了’，or ‘我先把上次借的书还了。’”

（刘珣等 2002,《新使用汉语课本》第2册 第42页）

In the 把 construction, the object is often something already known to both the speaker and the listener.

(Liu & Yao 2018, *Integrated Chinese*, 4<sup>th</sup> ed., Textbook 2, p. 142)





# 母语者语料: BCC

总字数约 150 亿字, 包括: 报刊 (20 亿)、文学 (30 亿)、微博 (30 亿)、科技 (30 亿)、综合 (10 亿) 和古汉语 (20 亿) 等多领域语料

BCC

把/p[一 两 三 四 几 数 十]

搜索

多领域

文学

报刊

对话

篇章检索

古汉语

搜索: 把/p[一 两 三 四 几 数 十];

共 56657 个结果, 这是第 1 - 100 条

高级设置: 上下文显示方式: 按上下文字数显示; 结果显示顺序: 顺序显示

统计

筛选

下载

高级

首页

上页

下页

末页

- |    |    |   |
|----|----|---|
| 1  | 全文 | 是哪位大神呢? 一年轻人, 觉得自己怀才不遇, 有位老人听了, 随即 <b>把一</b> 粒沙子扔在沙滩上, 说: “请把它找回来。”“这怎么可能!”接着老    |
| 2  | 全文 | 扔在沙滩上, 说: “请把它找回来。”“这怎么可能!”接着老人又 <b>把一</b> 颗珍珠扔到沙滩上, “那现在呢?”如果你只是沙滩中的一粒沙, 那你      |
| 3  | 全文 | 亮会变红? 月全食阶段月球完全位于地球本影内, 这时地球大气还会 <b>把一</b> 部分光折射到月球表面。在可见光中, 红光的波长较长, 蓝光和紫光       |
| 4  | 全文 | 你的需求! 童鞋们, 听听乔布斯年轻时每天凌晨四点起床,九点半前 <b>把一</b> 天工作做完.他说:自由从何而来?从自信来,而自信则是从自律来!        |
| 5  | 全文 | 艺片太文艺了, 不知道这些第五代。。。直到第N代导演为何偏偏要 <b>把一</b> 件简单的故事叙述成复杂的2B片。有3D所以我们当然不选择2B。         |
| 6  | 全文 | 不错的人。所以当你抱怨你对一段感情感到不值得的时候, 我会直接 <b>把一</b> 句话甩在你的脸上: 傻逼, 感情就是不问值得不值得。还是回丽江城里       |
| 7  | 全文 | 是很美妙。但找到一个好的生意模式确实有点可遇不可求。苹果这种 <b>把一</b> 个不好的生意模式愣是改成好的生意模式的情况恐怕真有点空前绝后。          |
| 8  | 全文 | 岁月是把杀猪刀, 有个被杀咯, 有个完全没事! 真的是一个对比啊我 <b>把数</b> 字证书删掉了…我脑残了…网银废了…天气真好, 出门游            |
| 9  | 全文 | 的好慢病推手这不是政治问题, 也不是医学问题, 是物理问题。我们 <b>把几</b> 亿年蕴藏在地壳之下的“碳”搬到大气中, 又用石油作为化肥和各种食       |
| 10 | 全文 | 。受伤了, 想倾诉, 却发现, 没有谁会是你的知音。只好小心翼翼的 <b>把一</b> 切尽藏在心里。假装很开心, 自己过得很好。开心了, 转身却发现他或     |
| 11 | 全文 | 会了, 已经不爱了。一学生迟到, 他向老师解释说: “我今天早晨要 <b>把一</b> 头猪送去配种。”老师问: “这事你爸爸不能做吗?” “不行, 一定要    |
| 12 | 全文 | 膝盖呈弯曲姿势。5.要把脚跟提起, 就是垫脚。6.人少时最好要 <b>把两</b> 臂展开把住扶手或电梯让生活更健康, 爱尼克斯净水机! 迷茫时看的8       |
| 13 | 全文 | 一样; 星期天的感觉就像快分手一样。大伯, 如果你有十亩地, 愿意 <b>把一</b> 半的收成献给党吗? 大伯: 愿意。记者: 两套房子呢? 大伯: 愿意。记  |
| 14 | 全文 | 轻人, 在你落败的时候绝不能读太多励志书籍, 读励志书籍有时候会 <b>把一</b> 个人读傻。因为此后你会相信你走的路一定是正确的, 即使偶遇失败,       |
| 15 | 全文 | 还可以睡上几小时; 10乔布斯年轻时每天凌晨四点起床, 九点半前 <b>把一</b> 天工作做完.他说:自由从何而来?从自信来, 而自信则是从自律来!       |
| 16 | 全文 | , 打磨你锐气, 消耗你的梦想, 拆散坚定过的爱情, 然后才能放心地 <b>把一</b> 个足够圆滑和现实的你, 送到社会, 婚姻, 现实里。大学, 多可怕, 多 |
| 17 | 全文 | 睡觉~~~我一直觉得我挺好的嘻嘻摩羯座的童鞋收。。。一夏利 <b>把一</b> 横穿马路的人撞飞了~我现在很恍惚~~吡饱了不想家用塑料袋装不卫           |
| 18 | 全文 | 想见到~洗个澡出门打球且。。。好像顺序弄反了哟~~神经兮兮的 <b>把三</b> 国杀卸掉掉~现在记不得密码了不是~我觉得大家都要疯啦!!! 爱吃         |
| 19 | 全文 | 稽啊, 好搞笑啊, 好可爱啊, ++!!!! 我受不了豆瓣了! 它居然 <b>把三</b> 枪拍案惊奇这种片子归类到了悬疑片里面!!! 这样好吗!!! 阿东妈   |
| 20 | 全文 | 孩我比较喜欢后面的蛋糕, 谢谢一妈表示完全不嫉妒, 米国二妈们是 <b>把几</b> 年的话一次性都说了出来你TX功力真是一流啊。。这都什么道德品质        |
| 21 | 全文 | 勿以恶小而为之, 所以你职责不愿意捐钱的人是善还是恶? 请你不要 <b>把一</b> 块钱都不愿意捐的事找任何理由来搪塞。就像外面下雨, 但有人还晒着       |
| 22 | 全文 | 点就睡觉, 看看表, 那个数字就是天方夜谭...似乎已经全乱套了...想 <b>把一</b> 个男人留在身边, 就要让他知道, 你随时可以离开他。哈哈, 吓人不? |



收藏



打印

# 母语者语料: [COCT 書面語語料庫2019](#)

COCT 書面語語料庫2019, query “把 P (一|兩|三|四|幾|數|十)” returned 13,325 matches in 10,807 different texts (in 319,712,694 words [168,090 texts]; frequency: 41.68 instances per million words)

No	File name	Solution 1 to 1000	Page 1 / 14
1	<a href="#">5/ba3b0932</a>	去留學，母親在機場哭得昏了過去，	<a href="#">把一</a> 個心肝寶貝送到那麼遠的地方真是肝腸寸斷
2	<a href="#">5/ba3b0932</a>	銷假上班。婚後第一次領到薪水時，他要求	<a href="#">把兩</a> 人薪水歸併在一起，由他負責計畫支配，因為
3	<a href="#">5/ba3b0932</a>	丈夫和她結婚以後，有一種「我終於	<a href="#">把一</a> 件想要的東西要到了」的心理。覺得
4	<a href="#">5/ba3b0932</a>	治療而獲得改善。海勒和海恩做過實驗，	<a href="#">把一</a> 群小貓放在圍欄內的儀器箱中
5	<a href="#">5/ba3b0932</a>	幻想的話，常能繼續讓孩童感到有趣，如	<a href="#">把一</a> 堆沙說是沙漠，一盆水說是海洋
6	<a href="#">5/ba3b0932</a>	遊戲，都能訓練眼和手的協調合作。	<a href="#">把一</a> 條或兩條毯子連接並摺疊起來，可做成
7	<a href="#">5/ba3b0932</a>	者。實驗的程序是這樣的，實驗者單獨	<a href="#">把十</a> 張人像的照片一一給受試者觀看，受
8	<a href="#">5/ba3b0932</a>	已經建立的實驗結果是否可以再度獲得證據。他	<a href="#">把十</a> 個實驗者分為兩組，羅生縮告訴一半的實驗
9	<a href="#">5/ba3b0932</a>	困難的字謎遊戲。所謂字謎遊戲就是	<a href="#">把一</a> 個英文字的字母顛倒打散，受試者必須把
10	<a href="#">5/ba3b0932</a>	Luchins 將一群人分成兩組，第一組他	<a href="#">把一</a> 個年輕人故意先說成是外向的，然後在同
11	<a href="#">5/ba3b0932</a>	比較容易取勝，就可考丙組，將來還是可以	<a href="#">把兩</a> 者合併起來，以電腦作業發展為多變事業；同樣地
12	<a href="#">5/ba3b0932</a>	為了她的「二齒」而大肆加工，	<a href="#">把一</a> 張嘴畫得像火盆，結果使人一
13	<a href="#">5/ba3b0932</a>	。例如在天花板上垂下兩根繩子，要他們	<a href="#">把兩</a> 根繩子繫在一起打個結。但是兩根繩子
14	<a href="#">5/ba3b0932</a>	關係的訓練」。我們剛才講過，創造是	<a href="#">把兩</a> 個完全沒有相關的東西拉在一起，所以我們可以
15	<a href="#">5/ba3b0932</a>	於是兩個就打起來了。這時候成人儘快	<a href="#">把兩</a> 人分開，然後重新替孩子解釋為什麼他們會打
16	<a href="#">5/ba3b0932</a>	自信。方法二：缺點消除法用記憶的方式，	<a href="#">把一</a> 、兩星期來所有消極的反應與思想都紀錄下來





# 母语者语料: COCT 口語語料庫2019

Your query "把\_P (一|兩)" returned 658 matches in 572 different texts (in 21,541,186 words [5,936 texts]; frequency: 30.55 instances per million words) [0.763 seconds - retrieved from cache]

<< >> >| Show Page: 1 Line View Show in random order New query Go!

No	Filename	Solution 1 to 658	Page 1 / 1
1	DA2014MYHBDShkd25J	經驗的也許他平常不太愛講話可是當我們	把一個熟悉的植物像艾草這個這樣熟悉的民俗
2	DA2014LSXFXD63J	講我們的課程大概一天到兩天就可以	把一套系統做好而且會讓大家更省成本其實也
3	DA2014LSXFXD63J	的時候告訴你大概三到六個月你只要	把一個水龍頭打開就好了就可以讓就其他沒
4	DA2014HYRYQD6JYSQ	工作如果我們就是做一個助理護士的訓練班	把一批一批的把他訓練出來然後我們再把這
5	DA2014HJMBSd3J	討論到這裡那等一下今天我要請你幫忙出任務	把一個包裹送出去同時記得把那個資料要帶回來
6	DA2014DAQJL48FBD139J	有很多土地不能夠蓋一個房子所以它就要	把一個腳幾個腳露出來這裡變成了我們吊腳樓
7	DA2014CHRWDTD20J	又低就想多賺一點錢就好了反正	把兩個小孩養大就可以了所以我現在我就不
8	DA2014YDFSHD13J	的時間管理安排事情的先後順序是的以及怎樣	把一個混亂的局面整理成一個非常可人的環境
9	DA2014YDFSHD13J	做一邊在展現同工的意義就比如說我教小朋友	把一個去骨的雞腿綁起來這個綁本身是我們想
10	DA2014SMHJCD2J	點不一樣的東西共同參與會有感情當我們	把一個可能還好的空間變成你今天12點以後你
11	DA2014ZTWZQD2J	場塔加單車賽中騎著菜籃車穿著普通的運動涼鞋	把一群裝備昂貴精良的男騎士遠遠拋在腦後
12	DA2014RWPXD15J	在花蓮我們靜思精舍要蓋寮房嘛那你可以	把一批志工都不會蓋房子的人然後把他們帶
13	DA2014RWPXD15J	文獻資料看到你一天帶領了所有的師兄就	把一間學校完全把它拆除而且分類得好又如何
14	DA2014BACCQD2J	年來到爽文國中十幾年來他和老師們努力	把一個學風低落的學校注入活力改頭換面早爽文
15	DA2014BACCQD2J	到現在為止我還覺得還滿有用的你想要	把一件事情做完你要付出很大的決心讓它
16	DA2014BZYJD14J	都會這樣握一下手握一下手所以我們	把兩隻手這樣子握起來它就是朋友很好的朋友
17	DA2014JSSXXLJZD20J	人就一定會認同才氣這一點真的是可以	把一個花瓶做得到無法想像的境界像最近我們
18	DA2014LSXFXD61J	的那個危害環境的所有的或者是說我們	把一個農作譬如說有機有機水梨有機水梨跟慣行
19	DA2014JXNDD3J	沒關係我們是剛開始那妳先做那大家都	把一個區塊一個區塊先設計好然後再來做磨合





# 无定宾语“把”字句的习得：看看中介语语料库

全球中介语语料库：已标注125,937,057字，85,278,546词

全球中介语语料库 (熟语料库) 导出数据清单-query "【把一】" 共86条						
内容	作文题目	hsk等	学习时	学习目	年龄	性别
【处】[zb除]除此之外，我们父女之间有一些固定习惯，那就他每天上班前【把一】颗糖放在我的枕头下，我一【】[Zq睁]开眼匆匆忙忙找糖。	《记我的》	5级		研究	27	W
我要【把一】件礼物送给妈妈。	看图作文 (母亲节送礼物, 3幅)				19	女
到家【的】[Zd]时丽丽【把一】【个】[zb份]礼物【】[Zq放]在她妈【】[Zq妈]手上，那一【刻】[Zc]她妈【】[Zq妈]很吃惊【】[Zq地]告诉她：“这条裙子这么漂亮，我很感谢你！”	《.....变了》					男
那天晚上，他【也】[zb又]偷偷地进【去】[Zd]写字的房间抄写【了】[Zd]，他一不小心【就】[Zc]【把一】本本子【】[Zq碰]掉下去，他有点儿害【怕】[Zc]被发现，但确定没有声音，没有动静，他才放心地继续写下去。	偷偷地帮爸爸					女
社交网络和公民义务两个月前我【把一】部讲艾未未生活的纪录片看完了。	社交网络和公民义务					
罗密欧要【知】[Zc]道真的情况【下】[Zd]因为【他】[zb她]【要】[zb会]【把一】[zb]封信【奇】[zb寄]给罗密欧【了】[Zd]。	罗密欧与朱丽叶的故事					
她【把一】大堆的玻璃杯子洗干【静】[zb净]后，一不小心就把它们都撞碎成碎片了。	易之思之					女
终于小偷被警察捕获，不过他不是先生也不是那个太太，这个事情【底】[zb后]，父亲笑着告诉我：“你下次来家，【把一】张票据贴在门上。	一件小事					
Page0001--	画蛇添足					
我们踢了一阵以后，我不小心【把一】位女同学的手机【打】[Zd]摔了。	好人				20-25	男
那时我在家里骑自行车玩儿，玩儿了一会儿我觉得很无聊，我【把一】张椅子，站在椅子起来，玩了一会儿，【把】[Zd]椅子倒下来，然后让我的腿抛光上了，我不知道怎么做，一直在哭，爸爸看见了就跑得把我。	难忘的一天					
小时候，爸爸【把一】只可爱的小猫给我养了。	我可爱的动物					
可是上了笔译课改变了想法，【把一】门语言翻译成别语言表达出来同样的意思，这样的想法比以前的更理解了翻译是什么。	一个学期	5级			40-49	M
晚上到了我想睡觉爷爷【把一】本故事【的】[Zd]书读给我一个好听的故事。	我的爷爷				20-25	M
爷爷带着小明去看奶奶，然后爷爷【把一】本故事书送给奶奶了。	看爷爷				20-25	M
到家【的】[Zd]时丽丽【把一】【个】[zb份]礼物【】[Zq放]在她妈【】[Zq妈]手上，那一【刻】[Zc]她妈【】[Zq妈]很吃惊【】[Zq地]告诉她：“这条裙子这么漂亮，我很感谢你！”	罗文芬的故事				15-19	M

# 无定宾语“把”字句的习得：看看中介语语料库

## COCT 中介語語料庫2019

← → ↻ 🔒 coct.naer.edu.tw/cqpweb/learners\_2019/concordance.php?th... 🔍 ☆ 🌐 ⚙️ J ⋮

[English]

Your query “把\_P (一|兩)” returned 15 matches in 14 different texts (in 704,270 words [1,963 texts]; frequency: 21.30 instances per million words) [0.864 seconds]

⏪ << >> >| Show Page: 1 Line View Show in random order New query ▾ Go!

No	Filename	Solution 1 to 15	Page 1 / 1
1	5dc5143f6b08c62109f50a08	在工作上管理或接觸的物質財富，就可以把一部分拿回家，沒有關係。反正是民眾的，	
2	5dc5143f6b08c62109f50a4d	判斷。我以前上班的時候沒有理財觀念，只把一部分的薪水〈存〉在銀行。我覺得錢	
3	5dc5143f6b08c62109f50af4	(煎蛋卷)跟英文的rice(飯)，日本人把兩種結合再一起，想出來的名字。蛋包飯是番茄醬	
4	5dc5143f6b08c62109f50afa	字幕一直看直到我懂意思。唱歌時我要把一首歌跟著音樂唱很多次，然後記住歌詞	
5	5dc5143f6b08c62109f50b43	內陷的顏色。先從下往上捲一半，把兩側的皮都往內折，最後再持續往	
6	5dc5144d6b08c62109f50b83	個用三D印刷來做的機器人，因此它把兩個新科技聯在一起。在法國創造它的工程師	
7	5dc5144d6b08c62109f50b9a	(2)多方嘗試並幫助他人。在這我把兩件事一起講，因為我相信他們是密切相關	
8	5dc5144d6b08c62109f50b9a	應該要造福社會；而我呢，則建議乾脆把兩者合而為一試試看，不錯吧！一個人事業	
9	5dc5146a6b08c62109f50ddd	等畢業了然後才去。到18歲政府會把一張通知寄給他們。他們先檢查身體然後沒有問題	
10	5dc5146a6b08c62109f50e29	的情況馬上就過去！後來罵他的老公，把一杯水丟到老公的身上。老婆一次哭	
11	5dc5146a6b08c62109f50e4b	累得走不了路了。姐姐，請妳幫我一杯牛奶拿進客廳來。姐姐說：「你真的	
12	5dc5146a6b08c62109f50f19	lt;燙&gt;到了。所以我也把兩杯冰水拿給他們。下午客服中心客服給我打電話	
13	5dc514876b08c62109f50fd5	早上到我們家來幫我們，還不要我們把一個月的房租付給他們。我們租房子的合約	
14	5dc514876b08c62109f510d2	到來。因為可以跟我小家庭出去放鬆心情，出門把一固禮拜的壓力釋放。雖然還要顧著小孩	
15	5dc514876b08c62109f5114d	鍋子裝著香噴噴湯底，想吃什麼都可以把一串一串的食物放進鍋里煮。沙爹&	

# “把”字宾语的有定与无定（储诚志 2021）

基本结论：

1. “把”字宾语并不都是有定的，有定、无定的规律要修正。
2. 教材教师强调必须有定，是误导学生。
3. 高级程度的学生在接受足够量的自然语言输以后，通过自然习得，还是能掌握使用无定宾语“把”字句，但不知道是否为教材、老师要求“有定”的教学多付了一些代价。





## 特别注意的几个问题：

1. 语法方面，不是语料库里有的就是对的： Google:

“大家听演讲，他可能跟你同意，可能不同意...”

“我经常性遇见跟你同意的问题，每次电脑管家修复一下就好了！”

—尽量使用加工过的熟语料库。

2. 分词问题很普遍

自动分词问题多，人工校对问题也多：

语言问题、标准问题、能力问题、态度问题、时间问题

3. 对字频表、词频表的不当解读。

1) 把累计频率/覆盖面等同于理解率：“只要学习500个汉字，就可以理解n%的中文文章”

【的确存在字词的“效用递减率”（周有光 1992），但是字词的文本覆盖率不等于读者的文本理解率。否则只要会一个“的”字，就能理解3-4%的文章内容？认识了“老”和“板”，并不能懂得“老板”的意思，】 《千字文》的教法不可行，也与此有关



## 网上开放的部分知名英语和ESL语料库

- British National Corpus (BNC): <http://corpus.byu.edu/bnc> (100 million words)
- **Corpus of Contemporary American English (COCA)**: <http://corpus.byu.edu/coca> (450 million words)
- International Corpus of English: <http://bit.ly/1vqnhY0> (1 million words)
- Corpus of Canadian English: <http://corpus.byu.edu/can> (50 million words)
  
- Michigan Academic Spoken English (MICASE): <http://bit.ly/1y4BEZa> (1.8 million words)
- British Academic Spoken English Corpus (BASE): <http://bit.ly/1vqmExG> (1.18 million words)
- British Academic Written English Corpus (BAWE): <http://bit.ly/1vqnNpc> (6.9 million words)
- Corpus of Global Web-Based English (GloWbE): <http://corpus.byu.edu/glowbe> (1.9 billion words)
- Corpus of American Soap Operas: <http://corpus.byu.edu/soap> (100 million words)
  
- Michigan Corpus of Upper-Level Student Papers (MICUSP): <http://micusp.elicorpora.info/>
  
- **International Corpus of Learner English (ICLE)**:  
<https://uclouvain.be/en/research-institutes/ilc/cecl/icle.html>



# 介绍几个中文和汉语中介语语料库

1. 北京语言大学BCC汉语语料库: <http://bcc.blcu.edu.cn/>
2. 语料库在线 (中国国家语委, 现汉、古汉、词表、工具)  
<http://corpus.zhonghuayuwen.org/index.aspx>
3. 北京大学现代汉语和古代汉语语料库: [http://ccl.pku.edu.cn:8080/ccl\\_corpus/](http://ccl.pku.edu.cn:8080/ccl_corpus/)
4. 中國哲學書電子化計劃: <https://ctext.org/zh>
5. --紅樓夢漢英平行語料庫 <http://corpus.usx.edu.cn/hongloumeng/index.asp>
6. 汉语国际教育动态语料库 (北师大, 汉语教材语料库):  
[http://www.aihanyu.org/basic\\_v2/index.html#/index?\\_k=ro7bp9](http://www.aihanyu.org/basic_v2/index.html#/index?_k=ro7bp9)
7. 全球汉语中介语语料库: <http://qqk.blcu.edu.cn>
8. HSK作文语料库: <http://hsk.blcu.edu.cn/>



# 介绍几个中文和汉语中介语语料库

台湾 國教院語料庫索引典系統: <https://coct.naer.edu.tw/cqpweb>

<a href="#">中研院平衡語料庫4.0</a>	<a href="#">COCT 口語語料庫2016</a>	<a href="#">COCT 口語語料庫2017</a>
<a href="#">COCT 口語語料庫2019</a>	<a href="#">COCT 中介語語料庫2016</a>	<a href="#">COCT 中介語語料庫2019</a>
<a href="#">COCT 書面語語料庫2016</a>	<a href="#">COCT 書面語語料庫2017</a>	<a href="#">COCT 書面語語料庫2019</a>
<a href="#">書面語語料庫(工作坊專用)</a>	<a href="#">COCT 書面語語料庫2015</a>	



# 语料加工和使用工具:

## 推荐上外:

<http://cascorpus.com/tool-list>

The screenshot shows the website [cascorpus.com/tool-list](http://cascorpus.com/tool-list). The page has a blue header with the site name and navigation links: 关于我们, 语料库资讯, 语料库知识, 常用语料库, 语料库工具, 研究文献, 语料分享, 注册/登录. Below the header is a navigation bar with tabs: 全部, 语料编辑工具, 语料平行工具, 语料标注工具, 语料检索工具, 其他语料库工具. The main content area lists several tools:

- SISU Aligner 1.0.0**: SISU Aligner 1.0.0是上海外国语大学语料库研究院开发的一款语料平行对齐软件, 可以实现“一对多”平行文本的对齐及编辑处理, 适用于双语/多语种语料库的平行对齐处理。 (485 views, 6月前)
- Antconc**: Antconc是由Laurence Anthony教授开发的一款单语语料库检索软件, 具有索引、词表生成、主题词计算、搭配和词簇提取等多种功能。 (120 views, 6月前)
- Pattern Builder**: Pattern Builder是梁茂成教授设计的一款赋码语料库检索辅助工具, 内嵌CLAWS4词性赋码集, 可以显示所选语言特征的正则表达式, 具有编写、调试正则表达式以辅助检索的功能。 (78 views, 6月前)
- EmEditor**: EmEditor 是一款Windows文本编辑器, 由日本江村软件公司 (Emursoft) 开发, 具有查找/替换字符串、分割/合并文件、多选区编辑等功能。EmEditor允许正则表达式使用, 支持超大数据文件处理, 支持支持 Unicode以及多文件的转码... (83 views, 6月前)
- NLPIR汉语分词系统**: 主要功能包括中文分词; 英文分词; 词性标注; 命名实体识别; 新词识别; 关键词提取; 支持用户专业词典与微博分析。 (176 views, 7月前)
- EditPlus**: (No description visible)

On the right side, there is a '热门工具' (Popular Tools) section with a similar list of tools:

- SISU Aligner 1.0.0**: (485 views, 6月前)
- NLPIR汉语分词系统**: (176 views, 7月前)
- Wordless**: Wordless是由上海外国语大学叶磊开发的一款免费开源语料库集成工具, 支持多语种、... (143 views, 6月前)
- Antconc**: (120 views, 6月前)
- ABBYY Aligner**: ABBYY Aligner是ABBYY旗下的一款双语对齐工具, 可以自动实现双语平行文本的对齐, ... (93 views, 6月前)





专为中文教师开发的工具平台：  
《中文助教 (ChineseTA™)》 储诚志 et al, 2014,2016)

《中文助教 (ChineseTA™)》是一个帮助中文老师编写、修改和评量教材和教学辅助材料的软件。该软件集成了很多功能。编写成套教材或准备临时性的教学辅助材料都涉及到很多费时费力而且容易出错的工作。在"中文助教"提供的环境中，很多事情都能迅速方便地完成。



# “中文助教”的功能

1. 课文加注拼音	11. 课文改换调序
2. 生词生字查找	12. 简繁体转换校对
3. 词表字表注释	13. 汉字注音校对
4. 汉字繁简对照	14. 词表字表项目的选择和排序
5. 字词分布索引	15. 已学教材字词的确定
6. 字词频率统计	16. 用户词库的维护
7. 生词密度和重现率标示	17. 字词频率和分布的分课统计与限定范围统计
8. 字词常用度标示	18. 字体调整
9. 新词旧词关联	19. 字词编辑、查找、替换
10. 词语随文注音翻译	等等



# Q & A



TCLT Tech-Talk Series  
@<http://www.tclt.us/talk/>

谢 谢



TCLT Tech-Talk Series  
@<http://www.tclt.us/talk/>