

# 语料酷！

*Corpora and online resources in the Mandarin classroom*

Simon Smith  
Coventry University

史尚明

Chinese Applied Corpus Linguistics  
Symposium, Lancaster University



5/1/2016

## Outline

1. *Corpus Linguistics in Chinese contexts*
2. *Corpus-based Chinese learning at XJTLU*
3. *Online Chinese learning at Coventry*

5/1/2016

Chinese Applied Corpus Linguistics  
Symposium, Lancaster University



## Definitions

- **Corpus**
  - 语料库
  - Searchable database of authentic language
- **Corpus linguistics**
  - 语料库语言学
  - Tools and approaches for searching a corpus

5/1/2016

Chinese Applied Corpus Linguistics  
Symposium, Lancaster University



## *Corpus Linguistics in Chinese contexts*

- 9 chapters by different authors
- 2 different “contexts”
- 6 on corpus analysis of English from China
  - e.g. Liang Maocheng, He Anping
- 3 Chinese corpus studies

5/1/2016

Chinese Applied Corpus Linguistics  
Symposium, Lancaster University



## Chinese corpus studies

- M Hoey & J Shao
  - Lexical priming theory
    - links corpus linguistics with psycholinguistics
    - applied to Chinese
- R Xiao
  - Passives and classifiers in Chinese and English
- A Kilgariff, S Smith, N Keng
  - Sketch Engine in Chinese learning

5/1/2016

Chinese Applied Corpus Linguistics  
Symposium, Lancaster University



## Classifiers in Chinese

- aka “measure words” 量词
- English sometimes uses classifiers
  - A sheet of paper
  - A piece of paper
  - A ream of paper
- But normally not
  - A bank
  - A gun
  - A car
- Chinese (almost) always uses classifiers

5/1/2016

Chinese Applied Corpus Linguistics  
Symposium, Lancaster University



## Classifiers in use

一	张	纸	Challenge for students!
yi	zhāng	zhǐ	
1	MW	paper	

A piece of paper

  

一	条	路	
yi	tiáo	lù	
1	MW	road	

A road

  

三	个	人	
sān	ge	rén	
3	MW	person	

3 people

Coventry University

## Please find usual measure word for X

X

Concordance - First query form - Mozilla Firefox

File Edit View History Bookmarks Tools Help

NUS - Centre for English Language C... x Concordance - First query form x Problem

http://beta.sketchengine.co.uk/auth/preloaded\_corpus/i-zh/ske/first\_form

Sketch Engine beta

user: Simon Smith corpus: Internet-ZH

Concordance Word List

Query Type: CQL

CQL: [tag="q"] "石头"

Make Concordance Clear All

Expert options: Query Type

## Please find usual measure word for X

X

吧。真奇怪。“阿婆却下狠心，命她立时，自己拿块石头坐了，从腰间掏出一方黄巾，擦了擦汗，还能不平安归来吗？如今他真的去了，阿婆心里倒像一块石头落了地；他不去，心却不在这；他去，索然无味，心就停止了一样，直到她说理想。我心想一块石头才落了地，不过她说我这么容易就想死。我说那好，那看就把电话挂了吧。这时我心里总算一块石头落了地，对方向口气凛然而肯定，完全听不过却头回的停不是一光线的“人工奇石”。“我这块石头是从中卫黄河宁夏的加回来的，少了4000... ? 带着疑问，记起决定那个问题。“小老乡，你的这块石头能卖吗？”记起问题。“当然能卖啊，... 记起为它打听了半年。“你是我老乡，我也不骗你，这块石头的成本不足一元钱，我每块卖上10元，我石头，为什么不把它挖走？”“你这么说？”“你说那石头哪？”从你说到时，就一直说到现在了，... 道都没到几分钟就把石头挖起来，看看大小，这块石头没有想像的那么大，都是被那个巨大的... 书中所说的那样多。那样详细。那山的山说，小到一块石头，一座亭子，一个山洞，都有它的不同... 看像一个小门角，也玩一玩那山说，这块石头，这块石头，这块石头，这块石头，这块石头，这块石头... 什么也没说。这说明了于是要拜拜了小鬼，好大一块石头，落地了，落地了，落地了，落地了，落地了... 只要用手在石上摸几下，便一切都明白了。原来，这块石头是比干的，心，没有被狐狸挖走，就飞... 变成了石头。而今比干都是忠臣，心魔相通。所以，这块石头很灵验。每回外出时，手下人只要在石... the Blarney Stone”（您早该来一下布拉克那说那块石头了）。原来在爱尔兰的Blarney城有块很... 的打，有人在山道上看见半块，么厚真重，就给了个石头片儿，国香半真面目个圈儿，过几天想... 最好听是这星说的“真稀奇”。我地呢了：心是一块石头落地，人心里地落了地。和几个... 的小女孩，那孩的用一种比大人还成熟的眼神望着两个石头一样洁白的老家伙，这是怎样的一种... 的打，[ ]每人5元。去呢超是16路，不过只有几个石头人，虽然还要10元门票很不懂这鬼... 象大就要吃果子，当时为雷小姐的婢女呢，那时拿一块石头把大能打死。通知大能死后魂不散，...

Page 1 of 99 Go Next Last

Chinese Applied Corpus Linguistics Symposium, Lancaster University

## Please find usual measure word for X

X

Frequency list

Frequency limit: 0 Set limit

Page 1 Go Next >

word	Freq
p/n 块石头	1517
p/n 个石头	140
p/n 颗石头	59
p/n 堆石头	45
p/n 种石头	21
p/n 座石头	19
p/n 些石头	16
p/n 条石头	13

Chinese Applied Corpus Linguistics Symposium, Lancaster University

## Two levels of collocation

- Inter-word, as English
  - ? Strong/powerful car/tea
  - ? 造成成果/后果/结果
- Intra-word
  - 结果/后果/成果/果然/如果
  - 水果/苹果/糖果/果酱/果冻
  - Level of analysis unavailable in English
- We want to teach:
  - What is possible?
  - What is frequent?

Chinese Applied Corpus Linguistics Symposium, Lancaster University

## Please find words that commonly go with 造成

造成 (0) Chinese GigaWord 2 Corpus: Mainland, simplified 频率 = 66956

SentObject	10214	19.8	Object	45691	3.7	Modifier	10282	2.9	Subject	26157	2.5	PP-以	34	1.0	PP-在	124	1.5
造成	255	9.14	损失	322	8.84	造成	400	10.13	灾害	581	8.39	造成	9	2.45			
造成	692	8.77	死亡	399	9.25	并	878	6.47	事故	555	8.07	造成	11	0.89			
造成	319	8.62	结果	351	9.07	未	395	8.4	影响	215	7.67	造成	34	1.0			
造成	278	8.5	人员	353	7.57	会	1042	8.33	后果	348	7.56	造成	10	0.07			
造成	159	8.29	平民	596	8.46	不会	234	8.0	原因	389	7.41	造成					
造成	240	8.27	损害	543	8.2	必须	78	7.84	火灾	157	7.39	造成					
造成	323	8.07	威胁	602	8.04	往往	75	7.58	洪水	237	7.38	造成					
造成	113	8.06	压力	545	7.9	甚至	117	7.54	事件	424	7.3	造成					
造成	278	8.0	冲击	353	7.57	日	1723	7.43	冲突	420	7.3	造成					
造成	77	7.63	事故	495	7.54	共	483	7.39	台风	142	7.23	造成					
造成	73	7.59	影响	833	7.46	当地	71	7.34	人体	172	7.22	造成					
造成	96	7.52	灾难	222	7.17	没有	677	7.09	内战	159	7.21	造成					
造成	66	7.5	人员	1495	7.1	一手	54	7.04	火灾	170	7.2	造成					
造成	641	7.45	水土	127	6.62	已经	237	6.6	洪灾	122	7.12	造成					
造成	186	7.33	伤害	350	6.56	造成	243	6.6	生态环境	186	7.04	造成					
造成	288	7.3	亏损	382	6.46	必然	44	6.56	车祸	108	6.96	造成					
造成	55	7.17	危害	165	6.45	所	208	6.56	打劫	129	6.72	造成					
造成	53	7.14	灾害	194	6.38	常常	36	6.53	失败	113	6.68	造成					
造成	45	6.56	负面	116	6.27	尚未	38	6.11	飓风	80	6.58	造成					
造成	32	6.22	国有资产	132	6.15	持续	36	5.88	环境	625	6.54	造成					
造成	39	6.19	巴勒斯坦人	132	6.13	也	317	5.86	金融危机	113	6.42	造成					
造成	23	6.19	士兵	138	6.09	最终	35	5.74	大火	82	6.4	造成					
造成	23	6.11	以色列人	88	5.92	想	77	5.61	大雨	65	6.21	造成					
造成	117	6.08	障碍	106	5.84	造成	18	5.53	北京奥运会	61	6.2	造成					

10 元旦当天，罗德岛州长刚上任两个小时就因该州的存款保险系统缺乏资金，而紧急宣布关闭该州的45家银行和信用社。**造成**不明情况的混乱。

11 这种生产体系严重诱发资源浪费和民族冲突的危险，从而对公民安全和国家的主权与完整**造成**难以估量的后果。

12 炸弹，一旦海湾地区爆发战争，必将**造成**严重后果。

13 在实际工作中，在某些特殊和某种情况下发生过由于偏差上造成的**现实**，**造成**了严重的后果。

14 他在谈到**造成**上述问题的原因时指出，1989年及此前几年的外贸顺差是在不断经济危机进行出口和严格限制进口使国家**造成**了不利的影响。

15 他警告说，危机的前景是可怕的，将会具有巨大的惯性 and 灾难，生态环境将进一步遭到破坏，在物质和道义上，广大人民群众要面临空前凶险的挑战，在参加活动中也不发生愉快的事情，要重视我国人民的传统美德，互相尊重，互相理解，互相学习，互相合作，在前进的道路上，也要不断排除，切实落实安全，崇尚文明诚信，在冷静、理智、宽容、理解、融洽地处理突发事件的进程中，使国家保持繁荣和稳定。

16 声明说，海湾战争正在面临着进一步扩大的危险，战争给该**地区**及**全世界**无法估计的**严重后果**。

17 文拉塔拉指出，海湾战争给国际和平、安全和世界经济**造成**严重后果，尤其是将严重影响到发展中国家。

18 他的家族因与邻居发生纠纷而伤害了对方，**造成**严重后果；他的侄儿染病去世，很有求人情，他都**造成**了严重后果。

19 对利用商业进行的各种非法牟利或进行错误活动，**造成**严重后果的，要予以刑事处罚或行政制裁。

20 严格遵守商业的各种非道德行为，损害商业信誉和道德规范，经营不善或发生其他非法行为**造成**严重后果社会危害的，要予以刑事处罚或行政制裁。

21 经海防炮响后，大部队四五百名队员追随我判官对外进行组织指挥，个别队员甚至临阵相倒，幸被及时抢救。

22 要严格落实保密制度，对违反保密规定并**造成**严重后果的，要严肃处理；对工作认真负责、成绩突出的单位和个人，同时，每个人在什么时候都能够应用最先进的技术，而又敢于献民资源或对环境**造成**严重后果。

23 其次由两院联合公布对使用国际人权公约的呼吁，由于双方不能建立正常关系，致使双方出现外交纠纷，尤其造成两院贸易关系而**造成**严重后果。

24 美国国务院发言人塔格特20日说，南斯拉夫的分裂很可能**造成**恶性后果，不仅影响南斯拉夫，还会**造成**整个欧洲混乱并损害联合国安理会的信誉，对**南斯拉夫**的分裂，不加以影响，国际局势，**造成**严重后果。

25 各国总统必须遵守联合国安理会的决议，对**南斯拉夫**的分裂，不加以影响，国际局势，**造成**严重后果。

# Please find frequent words which include character X

user: Simon Smith
corpus: Internet-ZH

[Concordance](#)

[Word List](#)

[Help on main menu](#)

---

[Help on Expert Options](#)

Expert options:

[Query Type](#)

[Character](#)

Query Type:

Character ▾

Character:

果

Make Concordance

Clear All

5/1/2016
Chinese Applied Corpus Linguistics Symposium, Lancaster University
14

[illegible]

# Please find frequent words which include character X

# Please identify the verb in these verb+object constructions (not SKL)

自訂語料庫搜尋範圍

Academia Sinica Balanced Corpus of Modern Chinese

使用說明 回首頁

指定語法結構範圍：

- \*全部
- ☐ 指定範圍：
  - ☐ 書面語
  - ☐ 演講稿
  - ☐ 劇本、台辭
  - ☐ 談話
  - ☐ 會議記錄

指定文體風格範圍：

- \*全部
- ☐ 指定範圍：
  - ☐ 記敘
  - ☐ 論議
  - ☐ 說明
  - ☐ 描寫

指定媒體形態範圍：

- \*全部
- ☐ 指定範圍：
  - ☐ 報紙
  - ☐ 一般雜誌
  - ☐ 學術期刊
  - ☐ 教科書
  - ☐ 工具書
  - ☐ 學術論著
  - ☐ 一般圖書
  - ☐ 其他
  - ☐ 視聽媒體
  - ☐ 會話談話

指定主題與學範圍：

- \*全部
- ☐ 指定範圍：
  - ☐ 哲學
  - ☐ 科學
  - ☐ 社會
  - ☐ 藝術
  - ☐ 生活
  - ☐ 文學

17

# Please identify the verb in these verb+object constructions (not 5分)

