

Chinese Text Project
中國哲學書電子化計劃
<http://ctext.org>

Donald Sturgeon (德龍)
djs@dsturgeon.net
香港城市大學

<http://ctext.org>

簡介

- 資料庫 + 數位化的研究工具
- 對象：中華地區、非中華地區的學者
- 動機：文本詮釋上的問題
 - 第一材料
 - 注釋
 - 翻譯
 - 相似段落
- 不依賴學校或機構的金錢支持而成立

起源和歷史

- 檢索先秦文獻的研究工具
- 範圍極小，較難的校勘、詮釋問題
 - 第一本：《墨經》
 - 如同“百科全書”
 - 並非文章、論文寫作方式；條目上下順序不重要
 - 原文中本身有明顯的類似“數據庫”的結構
 - 一段 = 一項資料 (record)
- 適用於一部原典的方法也適用於其它著作
 - 例如：《論語》的結構與《墨經》有相同之處

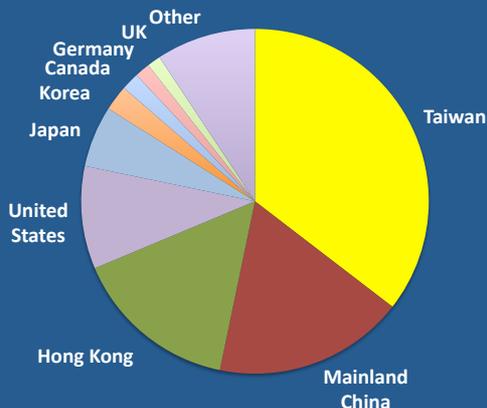
3

原典和數據

- 包括大多數先秦兩漢傳世文獻
 - 約2600萬字
- 傳世文獻影印資料
 - 約1300萬頁
- 可自行修改的維基資料
 - 約11億字
- 已註冊的自願貢獻者
 - 約6千人

4

網站訪客



- 人數
約25萬／月
- 身份
 - 研究者
 - 翻譯者
 - 學習語言人士
 - 大眾

5

證據基礎

子墨子言見漆絲者而歎曰漆於蒼則蒼漆於黃則黃所入者變其色亦變五入而已則為五色矣故漆不可不慎也非獨漆絲然也國亦有漆舜漆於許山伯陽禹漆於皋陶伯益湯漆於伊尹仲虺武王漆於太公周公

非獨染絲然也，國亦有染。舜染於許山、伯陽，禹染於皋陶、伯益，湯染於伊尹、仲虺，武王染於太公、周公。

非獨染絲然也，國亦有染。舜染於許山、伯陽，禹染於皋陶、伯益，湯染於伊尹、仲虺，武王染於太公、周公。

(Other references to "許由")

"This is true not only with silk dyeing; even a country changes its colour in response to its influences. Thus Shun came under the influences of Xu You and Bo Yang..."

(Relevantly similar instances of "然")

(instances of relevant grammatical constructions)

蒼則蒼染於黃則黃所以入者變其色亦變五入而以為五色矣一入故染不可不慎也非獨染絲也國亦有染舜染於許由伯陽禹染於皋陶伯益湯染於伊尹仲虺武王染於太公周公

6

兩種出版模式

- 傳統出版模式：
 - 預期（並要求）完善的數據
 - 1 製造完善的電子檔
 - 2 經專家校對製造完善的版本
 - 3 把內容引入數據系統
 - 4 完成
- “Web 2.0” 出版模式：
 - 預期數據含缺陷和錯誤
 - 1 把不完善的數據引入數據系統
 - 2 邀請使用者從中找出和糾正錯誤
 - 3 使用者自行加上標點符號、參與校勘
 - 4 第2、3步驟不斷重覆

9

自動收集資料

- 編輯字典難度高
 - 漢字數量頗高
 - 絕大部分的漢字是異體字、罕見字
- 收集第一材料為證據相當簡單
 - 足夠大的語料庫中，大多數漢字都會出現
 - 越是罕見字，出處越重要
 - 出處越多，數據越多

10

自動收集資料

徹         

U+5FB9 Seal script Jinwen Semantic variant Semantic variant Semantic variant Simplified character

Radical:	彳 + 12 strokes = 15 strokes total.	Kangxi Dictionary
References:	Guangyun: p.499#50 p.500#50 Kangxi: p.371#10 Cihai: p.519r6c01 GSR: 286.c Hanyu: v2,p0844#07	
Composition:	Component of: 劈 濼 礫	
Mandarin:	chè 彳 ㄓˋ	
Cantonese:	cit3	
Tang reconstruction:	djhiet tjet	Guangyun, commentaries
Shuowen:	《支部》徹：通也。从彳从支从育。 [2]	
Guangyun:	《廣韻·入聲·薛·徹》徹：通也，明也，道也，達也。又丑列切。 [2] 《廣韻·入聲·薛·出》徹：通也。 [2]	
Kangxi:	《康熙字典·彳部·十二》徹：〔古文〕徹徹《唐韻》《集韻》《韻會》从直列切，音徹。《說文》通也。《論語》齋徹乎。《註》周法，什一而稅，謂之徹。徹，通也，為天下之通法。又《廣韻》徹，達也。《左傳·成十六年》糞由基踵甲而射之，徹七札焉。《註》發達七札也。又《集韻》徹，道也。又刺取也。《詩·豳風》徹彼桑土。又治也。《詩·大雅》徹田爲糧。《傳》徹，治也。又去也。《禮·士冠禮》徹筵席。《左傳·宣十二年》軍衛不徹警也。《周禮·天官·膳夫》卒食以樂，徹于寤。《禮》天子食終徹器之時，作樂以徹之。又發也。《詩·小雅》徹我牆屋。《說》徹發我牆屋也。又《揚子·方言》徹，列也。《釋名·獨斷》羣臣異姓有功封者，稱曰徹侯，避武帝諱，改曰通侯，或曰列侯也。又《唐韻》丑列切《集韻》《正韻》蔡列切，从巛入聲。《說文》《集韻》或作《足駁》。 [2]	
Fanqie:	直列（《四書章句集注·論語集注·八佾第三》）直列（《四書章句集注·孟子集注·公孫丑章句上》）蔡列（《四書章句集注·孟子集注·滕文公章句上》）直列（《廣韻·入聲·薛·徹》）丑列（《廣韻·入聲·薛·徹》）丑列（《廣韻·入聲·薛·出》）	
Unihan definition:	penetrate, pervade; penetrating	

11

自動收集資料

Example usage

- 《論語·八佾》：三家者以雍徹。 [\[2\]](#)
The three families used the Yong ode, while the vessels were being **removed**, at the conclusion of the sacrifice.
- 《孟子·離婁上》：將徹，必請所與。 [\[2\]](#)
And when they were being **removed**, he would ask respectfully to whom he should give what was left.
- 《禮記·曲禮上》：客徹重席，主人固辭。 [\[2\]](#)
When the visitor (wishes to) **remove** one or more, the host should firmly decline to permit him to do so.
- 《莊子·大宗師》：已外生矣，而後能朝徹。 [\[2\]](#)
This accomplished, his mind was afterwards clear as the morning.
- 《道德經》：有德司契，無德司徹。 [\[2\]](#)
(So), he who has the attributes (of the Dao) regards (only) the conditions of the engagement, while he who has not those attributes regards only the conditions favourable to himself.
- 《商君書·兵守》：客至而作土以為險阻及柞格阱陷，發梁徹屋，給徒，徙之。 [\[2\]](#)
So as to make, at the approach of the invaders, earthworks as an obstruction, and traps, chevaux-de-frise and pitfalls, to pull down the supporting beams and to tear down the houses, to transport what is transportable.

12

字符識別

- 字符識別 (OCR)
 - 可直接應用到傳世文獻的第一材料
 - 效果不完善，會出現錯誤
 - 仍然可以提供有用的資料
 - 例如，谷歌圖書：雖然OCR效果不佳，可以實現中文傳世文獻的檢索功能
 - 可以透過OCR結果把電子版與影印本相連結
 - 可快速在第一材料中進行檢索
 - 可快速驗證電子版的正確性

15

字符識別

越庶士御事又曰義爾邦君越爾多士尹氏御事又曰肆哉爾庶邦君越爾御事是也者周書越大字語與一篇字而外義

越猶及也書召誥曰惟四月既望越六日乙未言自既望及乙未六日也下文曰惟丙午臚越三日戊申亦謂自丙午及戊申三日也後皆放此

越若亦及也召誥曰越若來三月來至也見爾言及至三月也越若來三月何惟丙午臚為句某氏讀越下若來為句傳曰於順來火之說見經義述聞下

文曰若翼日乙卯又曰越翼日戊午是越與若皆及也若與及本同義連言之則曰越若矣漢書律厯志引武

越庶士御事又曰義爾邦君越爾多士尹氏御事又曰肆哉爾庶邦君越爾御事是也者周書越大字語與一篇字而外義

越猶及也書召誥曰惟四月既望越六日乙未言自既望及乙未六日也下文曰惟丙午臚越三日戊申亦謂自丙午及戊申三日也後皆放此

越若亦及也召誥曰越若來三月來至也見爾言及至三月也越若來三月何惟丙午臚為句某氏讀越下若來為句傳曰於順來火之說見經義述聞下

文曰若翼日乙卯又曰越翼日戊午是越與若皆及也若與及本同義連言之則曰越若矣漢書律厯志引武

詞一

三

[View](#) [Edit](#) [Editing help](#)

16

未來展望

- 增加參與者
 - 越多參與者 => 越多編輯者
- 新增原典、注釋本、翻譯本
 - 更多資料 => 更好的自動化結果
- 廣泛使用自動化、提供統計
 - 電腦使得新的方法變為可能