战国词汇发展的时序以及在楚简和其他先秦典籍研究中的应用[1]

陶红印*

内容提要：本文尝试说明如何借助于电子语料库及语料库工具来确立战国文献词汇发展的时序性。我们主要利用语料库语言学中的一致性分析工具及逐字索引和搭配研究等方法把语词在若干战国文本中的分布精确地显示出来，进而在此基础上描绘楚竹简文本的语言特征，并用于判断其他先秦古籍的时代特征。我们的考察说明，这种研究方法完全可以扩大到其他文本研究工作中去，特别是涉及多种文本比较的工作。语料库分析的结果显示，具有关键字和关键词语价值的成分可以是大范围、多类型的，不应该限于个别语词，更不限于名词性成分的范围。同时，以统计为基础的语言用法作为文本比较的参考项，应该有一定程度的科学客观性，可以帮助我们判断古籍的时代特征。同时我们也强调，语言统计方法有其局限性，因此我们必须力争结合多种类型的证据实现对文献的更为科学的判断。

关键词：战国词汇、语料库、一致性分析、逐字索引、搭配研究、楚简

一、 战国文本关键词的确认与时代判断

利用语言特征确立文本的历史年代，古今中外的学者早已有过大量尝试。不过所用方法不尽相同。在近代以来比较有系统的论述者中，瑞典汉学家高本汉(Bernhard Karlgren)在上个世纪三十年代就尝试利用语法结构特征对《左传》及其他先秦文献进行断代分析。后来加拿大学者杜百胜(W. A. C. H. Dobson)以及最近美国学者白牧之(Bruce Brooks)等人也有类似的尝试。不过仅仅利用语法特征来判断文本时代有一定的局限性。以色列汉学家尤锐(Yuri Pines)曾经对此有过评论，[2]认为古代的典籍作伪者可能对语法特征有更清楚的意识，因此他们往往能利用语法特点模仿古书。因此他比较倾向于采用词汇分析的方法。刘笑敢从他的博士论文起就在一系列著述中利用综合的方法来判断老庄文献的时代问题。[3]他所用到的方法包

*洛杉矶加州大学亚洲语言与文化系及应用语言学系教授。（电邮：tao@humnet.ucla.edu）

[1] 本项研究计划为香港中文大学刘笑敢教授所主持的香港政府大学教育资助委员会资助的项目「《老子》文本两千年之演变及相关简帛资料研究」之一部分，并受到洛杉矶加州大学「数字人文计划」的资助。

[2] Pines, Yuri. “Lexical Changes in Zhanguo Texts”， Journal of the American Oriental Society, 122.4（2002）: 691-705。

[3] 见刘笑敢: 《庄子哲学及其演变》 (北京：中国社会科学出版社, 1988)。刘笑敢:《〈老子〉：年代新考与思想新诠》（台北：东大图书公司，1997）。Liu, Xiaogan. “From bamboo slips to received

括句式结构、词语变化、音韵模式以及文体风格等。为庄子内外篇时代的确立和了解《老子》文本在两千年里的演变规律提供了一个有效的研究范式。当然这里只是略举数例，类似的研究成果还有很多，这些累积的成果为古典文献的断代和深入理解提供了坚实的基础。不过到目前为止我们还很少见到利用电子语料库手段研究文本时代特征的实例（日本学者山田崇仁的一系列研究可看作少数例外[4]）。在电子文本处理技术已经比较成熟的今天，这不能不说是一个遗憾。利用电子手段研究文本的最大好处是穷尽性和精密性。如果文本材料可信、研究者能够确立有意义的检索项目，利用电子检索手段可以在很短的时间内将所有满足条件的相关语料搜出，并且作出不同类型的排比和分类，便于研究者探索其中的规律。[5]我们这里试图在这方面做一些初步的尝试。本文的主要目的是：1）确立战国文献词汇发展时序性。这方面的工作我们将基于较大规模的语料并借助于语料库工具来操作，希望能为今后类似的研究提供方法论上的尝试；2）在此基础上描绘楚竹简文本语言特征以及判断其他先秦古籍的时代特征。

如上所述，尤锐强调研究词汇分布和变化规律对古籍断代工作的价值。他所提出的研究战国文本词汇变化的具体方法包括下属内容。第一步是选择具有相对明确历史时代特征的文本，把这些文本所代表的时代作为基本时段参考点。他所选择的八个文本及分期如下：

春秋末期/战国早期（公元前 5 世纪或更早）：《左传》、《论语》

战国前期（公元前 5 世纪末至 4 世纪）：《墨子》

战国中期（公元前 4 世纪末）：《孟子》、郭店楚简

战国晚期（公元前 3 世纪末或更晚）：《荀子》、《韩非子》、《吕氏春秋》

第二，选择一组在早期文本中少见或未见而在晚期文献中较为常见的单字或词汇组合，把这些语词成分放在一起作为判断文本时代的参考项。他所选择的语汇包括以下几个：（1）“弩”以及“机、枢”（这一组语汇表示弓箭及发射装置）。据他

---

versions: Common features in the transformation of the Laozi，” Harvard Journal of Asiatic Studies 53.2（2003）：337-382。刘笑敢：《老子古今：五种对勘与析评引论（上、下卷）》（北京: 中国社会科学出版社, 2006）。

[4] 见山田崇仁：「『國語』韋昭注引系譜資料について－N-gram 統計解析法による分析－」『立命館史学』（2001）22 号。山田崇仁：「電脳東洋学的授業の試み 授業の実践報告」『漢字文献情報処理研究』（2002）第 3 号，漢字文献情報処理研究会。山田崇仁：「歴史記録としての『春秋』－N-gram モデルと統計解析法による分析－」『中國古代史論叢』（2004a）立命館東洋史学会叢書二。山田崇仁：「『孟子』の成書時期について－N-gram と統計的手法を利用した分析－」『立命館東洋史学』（2004b）第 27 号。山田崇仁：「N-gram モデルを利用して先秦文献の成書時期を探る－『孫子』十三篇を事例として－」，東京大学東洋文化研究所附属東洋学研究情報センター「アジア研究情報 Gateway」（2004c）。山田崇仁：「中国戦国期の語彙量について－N-gram とユールの K 特性値を利用した分析－」『漢字文献情報処理研究』（2004d）第 5 号。山田崇仁：「『礼記』中庸篇の成書時期について－N-gram モデルを利用した分析－」，『中国古代史論叢』続集。立命館東洋史学会叢書四（2005）。山田崇仁：「『周禮』の成書時期・地域について」『中国古代史論叢』（2006）三集。立命館東洋史学会叢書五。感谢佐藤将之先生提醒我注意到山田崇仁所做的相关研究。

[5] 见洪华清、陶红印（稿）：《老子》五种版本网络检索版及相关问题（香港中文大学中国哲学文化研究中心）。

考察，这组词汇在《孟子》以前的早期著作中不见，但在晚期著作中常见，特别是在晚期著作中它们既有字面意义的用法也有引伸意义的用法；（2）“仁义”，这两个单字的使用在早期著作中已有，但是作为有组合意义的合成词则见于《墨子》及后期文献；（3）“万物”，其中“物”字表示抽象的意义、“万物”表示不同事物的用法在《墨子》和《孟子》中偶见，但在郭店楚简（《老子》和《太一生水》）及后期文献中比较普遍。（4）“万乘”，这个说法只是在《墨子·非攻》及以后的文献中才能见到。（5）“理”，表示抽象的秩序、原则、机制的意义出现在战国中晚期。（6）“阴阳”，在战国早期文献中偶尔出现，多表示具体的自然物质现象，真正表示哲学意义上的元素对立出现在战国晚期（公元前三世纪以后），尤其是《吕氏春秋》中。（7）“布衣”，早期文献中的用法主要表示一种普通衣料，真正表示地位较低的普通人的用法出现在公元前三世纪以后的战国晚期文献中。

在确立了断代文本和关键词以后，尤氏以它们为标准来考察写作时代有争议的两个文本：《孙子》和《商君书》。就《孙子》而言，尤氏认为这部作品的用词现象比较接近《墨子》和郭店楚简，应属于前四世纪中期左右的作品。其证据是，《孙子》中有“弩、机”的用法，有一次“仁义”的用法，但没有“万乘”和“万物”的说法，“理”出现两次，没有“布衣”的用法。有阴和阳的提法但不太具有哲学上的意义。也就是说，《孙子》一书中能发现的战国后期语言的特征不多。用同样的办法考察《商君书》，尤氏认为这部作品更接近《孟子》和郭店楚简而跟战国后期文献很不相同。

撇开这里提到的相关参考文献的时代特征的可信度不论，尤氏的这种研究方法的优势是把时代特征和语言特征结合起来考察，而且所用到的词汇标准是成套和多项的，在一定程度上避免了个别词语的偶然因素。当然他的做法也是有不少问题的。主要局限性似乎体现在下面这些方面。首先，他所用到的大多数数据很可能是手工统计的结果，有其不精确性及范围的有限性。例如作者统计的《墨子》和我们的检索结果有一定的出入：“仁义”尤氏列出 19 例，我们发现有 27 例；万物，列出 2 例，我们发现有 5 例；理，尤氏列出 1 例，我们认为至少有 6 例。另外一个缺陷是，尤氏选取的关键词的词类类型有限，几乎全部为名词性概念。这类语词固然有用，但考察范围不一定要局限在这一类上。我们后面的结果将显示，非名词类的语词的用法有时在时代分布上可能有更明显的区别（也可以参看山田崇仁所用的 N 元语法的方法）。另外，虽然这里作者选用了一组词语，但毕竟数目还是有限的，因而个别偶然因素还是不能完全排除（例如表弓箭类器物的名词在不同类型文本中出现的几率对文本断代可能有一定的影响）。为避免此类错误，我们认为应该充分利用语料库手段的优势，扩大参考词项的范围。最后，尤氏所选用的具体词语基本上是依靠个人语感。语感的好处是，它通常是基于作者长期研究积累的结果，有坚实的研究基础。但缺点是有可能带着研究者的偏见。我们认为语感的运用应该和一些比较客观的手段结合起来，这样才能更为全面、可靠。在这方面语料库手段应该能够为我们提供帮助。而这也是我们撰写本文的出发点。

二、语料库分析手段及文本的选择

常见语料库分析手段包括逐字（逐词）索引（concordance）、词语列表（wordlist, 对语料库中所有语词穷尽列出）、关键词表（keyword list, 即通过对参考语料的对比，按照统计规律求出研究语料中的非同寻常的词语）、N 元语法（N-Gram,即任意大小的符号串的排列及统计）等。不过对我们当前的目的来说，比较相关的还是一致性分析法（Consistency Analysis）。这个分析手段的实质是把语料库中每个文本所用到的语词首先单独列表，然后排列比较各词表的异同，从中求出全部频率和文本覆盖范围。[6]我们这里所用的一致性分析工具为 WordSmith Tools 中词表（wordlist）主工具组下的一个子程序。[7]

由于目前还没有以分词为基础的电子文本，我们只能用以单字为基础分析文本材料。在单字的统计基础上再使用逐字索引的方法，求得常见的语词组合。这种确认词语的做法虽然比较间接，但优势是比较客观，因为它主要是以组合单位的统计数据为基础的。

我们所用的语料库以尤锐所列为基础，再加上上海博物馆所出的战国楚竹书前四册的内容。[8] 也就是说，从时代顺序上来看，第一组包括《左传》、《论语》；第二组：《墨子》；第三组：《孟子》（及竹简文献）；第四组：《荀子》、《韩非子》、《吕氏春秋》。因为本文的目的是以传世文献作为参考语料库，以楚竹书为比较对象，所以在比较语词分布的时候我们首先排除楚简文本。具体操作程序为，首先得出传世文献中的单字（语词）的一致性分布表，然后在这个基础上确认一批关键字（语词），再用这批关键字考察它们在竹简文本中的用法，最后进一步跟传世文本进行比较并运用到其他先秦古籍的断代考察上。

三、战国文本语词的一致性分布及关键词表

首先把传世文献的《左传》、《论语》、《墨子》、《孟子》、《荀子》、《韩非子》和《吕氏春秋》七个文本依顺序排列，用 WordSmith Tools 中的一致性分析工具求单字（或语词）分布，可得到如下几组结果（样本见附录）。

1）在所有七个文本中出现的字；

2）只在六个文本中出现的字；

---

[6] 由于语词的分布范围区域是这类工具的主要考察对象，有的语料库工具就直接用“语域”（Range）命名（见 Heatley, A., Nation, I.S.P. and Coxhead, A. 2002. RANGE and FREQUENCY programs. http://www.vuw.ac.nz/lals/staff/Paul_Nation.）。

[7] 见 Scott, Mike. WordSmith Tools. (Oxford: Oxford University Press，2008).

[8] 《上海博物馆藏战国楚竹书》篇目。第一册（三篇）：《孔子诗论》、《缁衣》、《性情论》。共 98 简。第二册（七篇）：《民之父母》、《子羔》、《鲁邦大旱》、《从政・甲》、《从政・乙》、《昔者君老》、《容成氏》。共 116 简。第三册（四篇）；《周易》、《中弓》、《恒先》、《彭祖》。共 108 简。第四册（九篇）；《采风曲目》、《逸诗・交交鸣鸎》、《逸诗・多薪》、《昭王毁室昭王与龚之月隼》、《柬大王泊旱》、《内豊》、《相邦之道》、《曹沫之陈》。共 125 简。

3）只在五个文本中出现的字；
4）只在四个文本中出现的字；
5）只在三个文本中出现的字；
6）只在两个文本中出现的字；
7）只在一个文本中出现的字；

考察发现，两头比较极端的结果除了若干个条目价值较大外，大多不十分有用。这些极端的结果包括在所有七个文本中出现的字和只在一、两个文本中出现的字，。它们的有限性在于，能在所有七个文本中出现的字多为常见词，典型的例子包括之、其、者、也等；只有个别非虚词用法比较有意义，特别是那些呈现出从战国早期到晚期递升趋势的非虚词性成分。而另一种极端的情况，即只在一两个文献中出现的字，它们大多为特殊词语和专有名词。例如《左传》中的地名、人名。大多数比较有意义的单字是能在六到三个文本中出现的。如果发现一个字能在六到三个文本中出现，我们就进一步寻找那些在早期文献中少见、未见而在中晚期常见的字。这种递升的分布模式可以说明这些字（或相关语词）是比较后起的。

下面显示的是我们根据一致性分析工具的原始数据（样本见附录）[9]而进一步确认的几组关键字及其在其中传世文献中的分布情况。这里总共有 44 个单字，每个单字一般都有依时代而递增的发展趋势，特别是晚期递增的趋势相当明显。

[9] 我们这里没有采用正态化（normalized）数据，主要原因是有很大一部分字数频率很低，而且一般上一致分析也没有利用正态化数据。

表一、在六到七个文本中出现的重要字项

单字	频率	文本	左传	论语	墨子	孟子	荀子	韩非	吕氏
論	224	7	1	2	7	2	57	68	87
謹	96	7	2	3	9	2	45	16	19
巧	116	7	2	5	13	4	20	37	35
變	219	7	4	7	6	15	60	58	69
避	86	6	1	0	17	4	11	31	22
參	85	6	1	3	3	0	20	47	11
誠	161	6	1	2	0	21	66	34	37
愚	199	6	1	9	32	0	60	65	32
化	165	6	2	0	1	4	74	16	68
靜	111	6	2	1	3	0	22	35	48
審	131	6	2	1	12	0	24	23	69
繩	55	6	2	0	2	2	19	16	14
妄	55	6	2	0	1	1	10	32	9
削	85	6	2	0	2	4	12	47	18
嚴	92	6	2	0	3	3	14	48	22
術	295	6	3	0	18	5	50	166	53
廷	54	6	3	1	0	2	12	26	10
形	167	6	3	0	7	4	55	46	52
精	119	6	4	1	2	0	25	22	65
決	88	6	5	0	2	4	15	32	30
理	293	6	5	0	9	7	105	86	81

注：“频率”指全部文献中的出现次数；“文本”指的是在几个文本中出现。

表 2、在三到五个文本中出现的重要字项

单字	频率	文本	左传	论语	墨子	孟子	荀子	韩非	吕氏
俗	159	5	0	0	14	5	79	15	46
勢	251	5	0	0	1	5	27	175	43
肖	173	5	0	0	18	5	26	45	79
全	95	5	0	0	7	1	17	23	47
智	472	5	0	0	46	32	9	240	145
境	79	5	0	0	12	3	8	34	22
拙	36	5	0	0	1	2	5	17	11
圓	38	5	0	0	1	1	4	11	21
悲	37	5	0	1	1	0	2	8	25
擅	70	4	3	0	0	0	20	43	4
雜	36	4	0	0	5	0	19	7	5
僻	28	4	0	0	3	0	11	8	6
蹠	38	4	0	0	0	4	11	16	7
恣	32	4	0	0	0	1	9	8	14
樸	28	4	0	0	4	0	8	10	6
卻	41	4	0	0	0	5	3	13	20
真	27	4	0	0	1	0	3	10	13
詘	33	3	0	0	0	0	17	6	10
驗	28	3	0	0	0	0	6	10	12
誹	30	3	0	0	0	0	3	9	18
虜	30	3	0	0	0	0	3	12	15
諭	25	3	0	0	0	0	3	5	17
暗	22	3	0	0	0	0	3	13	6

注：“频率”指全部文献中的出现次数；“文本”指的是在几个文本中出现。

由于表中所列项目是以字为单位，这些单字跟词语的关系不是一一对应的，也就是说，有的可能独立成词，有的可能只是词语的一部分，或者两种情况并存。下一步的工作就是通过单字寻找语汇，包括词或类似词的单位。我们任意规定如下条件作为选定关键词语的标准：1）构成成分只限于两个单字的范围；2）与关键字组合的成份一般不是虚词类成分（如“之乎者也”等）；3）组合成份整体结合在一起的出现频率一般不少于 5 次。注意这里所说的“词语”不一定能严格地对应词典上的定义，只要字的组合达到规定的频率标准我们就认为它们有常见性。

以下是我们找到的 5 笔以上的组合成份及其频率（X 代表关键字；数字表示出现次数；-表示缺少能满足条件的实例）。

（单字在 7 － 6 个文本中出现）

(7)
論: 此 X 13
謹: 不 X 7; 必 X 5
巧: X 言 8; X 拙 6; X 匠 5
變: 應 X 8; X 化 9; 萬 X 8; X 法 6; X 更 5; X 數 5
(6)
避: X 逃 9
參: X 伍 8; 名 X 5; X 驗 5
誠: X 信 12; 者 X 9; X 則 7; X 能 6; X 義 5
愚: X 者 33; 欺 X 10; X 智 10; X 人 9; 臣 X 7; 惑 X 7; X 眾 6
化: 知 X 10; 變 X 10
靜: X 郭 20; 動 X 13;
審: 必 X 5
繩: X 墨 12; 中 X 7;
妄: X 舉 7; X 發 6
削: X 地 9;
嚴: X 刑 16; X 罰 5; 威 X 5
術: 法 X 36; 道 X 10; X 行 5; 心 X 6; 用 X 5
廷: 朝 X 21; X 理 6
形: 無 X 11; X 體 9; 地 X 6; X 埶 5; 有 X 5
精: X 神 14; X 氣 8;
決: 能 X 9
理: X 義 9; 文 X 16; 道 X 13; 行 X 9; 勝 X 6; 事 X 6; 成 X 6; 循 X 6; 腠 X 6; 當 X 6; 廷 X 6; 大 X 6; X 足 5; X 督 5

（单字在 5 － 3 个文本中出现）
(5)
俗: 世 X 26; 習 X 9; 成 X 6; X 主 7; 風 X 5; 殊 X 5; X 儒 5
勢: X 重 13; 權 X 10; 威 X 7; 處 X 6; 得 X 5; X 亂 6
肖: 不 X 173; X 人 10; X 主 8
全: 能 X 5
智: X 者 28; X 伯 26; X 能 23; 為 X 12; 竭 X 8; X 士 11; 愚 X 10; X 過 8; 賢 X 6; 辯 X 6; X 慮 7; 事 X 7; 自 X 6; X 氏 6; X 短 5; X 慧 5; 去 X 5; X 力 5; 任 X 5
境: X 內 32; 四 X 12; 邊 X 10;
拙: 巧 X 6; 為 X 5
圓: 方 X 14; X 道 8
悲: 不 X 10; X 哉 9
(4)
擅: X 為 6; X 行 5; 得 X 5; X 作 5
雜: (學 X 4)
僻: (X 陋 4)
蹠: 盜 X 20; 桀 X 8

恣:　X 睢 6; 驕 X 5;
樸:　-
卻:　-
真:　-
(3)
詘:　田 X 5
驗:　參 X 5
誹:　X 謗 8
虜:　-
諭:　-
暗:　-

下一步的工作就是拿这些靠统计数字而确认的关键字及其组合来考察它们在楚竹书里的用法，并跟传世文献加以比较。由于篇幅所限，我们在这里不可能详细讨论每一个用例。作为方法上的演示，我们下面只在每一组中挑出一个关键字词来考察。这里被挑选的字的共同点是组合格式较多（即在组词方面比较有能产性）。

四、若干关键词语在传世文献和楚简中的分布

4．1．巧

单字	频率	文本	左传	论语	墨子	孟子	荀子	韩非	吕氏
巧	116	7	2	5	13	4	20	37	35

A) 单字用法：上博 4 次；郭店：5 次；可见竹书用法接近《论语》。
B) 固定搭配: X 言 8; X 拙 6; X 匠 5
　　巧言 8：《左传》1，《论语》1；上博 2，郭店 1。
　　巧拙 6：《荀子》5，《吕氏春秋》1；竹书：无
　　巧匠 5：《韩非子》1，《吕氏春秋》4；竹书：无
C) 结论: 整体上看竹书“巧”字用法接近战国早期文献。

逐字索引　　（S=上博, G=郭店）

忻>之志則曼■。人之〔攷〕**<巧>言** 利　<詞>者，不又<有>夫	S
夫亙<恆>怡之志則縵。人之攷**<巧>言** 利　<詞>者，不又<有>	S
也，不有夫恒怡之志則縵。人之**巧言** 利詞者，不有夫詘詘之心則	G
>，進谷<欲>孫<遜>而**毋攷**<**巧**>，退谷<欲>　〈？〉而毋亟	S
不　<食>若才<災>，腹中攷<**巧**>卜　日　又<變>，古<故>父	S
，民利百○（倍）。○（絕）○（**巧**）棄利，○（盜）惻（賊）亡又	G
其甬（用）不○（窮）。大○（**巧**）若○（拙），大成若詘，大植	G
），怒欲盈而毋希，進欲遜而**毋巧**，退欲循而毋輕，欲皆度而毋	G

成孝。君子孝子，不食若才腹中，**巧變**故父母安……是謂君　　G

4．2．　誠

单字	频率	文本	左传	论语	墨子	孟子	荀子	韩非	吕氏
誠	161	6	1	2	0	21	66	34	37

A) 单字用法：上博 4 次；郭店：4 次；超出左、论、墨，少于荀、韩、吕。
B) 固定搭配: X 信 12; X 能 6; X 義 5
　　誠信 12：《孟子》1，《荀子》5，《韩非子》3，《吕氏春秋》3；竹书：无
　　誠能 6：《荀子》3，《吕氏春秋》3；竹书：无
　　誠義 5：《荀子》2，《吕氏春秋》3；竹书：无
C) 结论: 整体上竹书接近战国中期文献。

逐字索引

民孰弗從？形於中，發於色，其**誠**也固矣，民孰弗信？是以上之恒　　G
顏子、子路之事孔子也。事之者，**誠**事之也。知而弗事，未可謂尊賢　　G
天下。仁覆四海，義襄天下，而**誠**由中心行之，亦君子已！　　G
，商湯之舉伊尹也。舉之也者，**誠**舉之也。知而弗舉，未可謂尊賢。　　G
俊、舜>之　<德><丌其>城<**誠**>　<賢>矣，　<播>者<諸>　　S
又<有>命自天，命此文王，城<**誠**>命之也■，信矣■。孔＝<孔子　　S
<　、俊、舜>之　<德>則城<**誠**>善亦　。先王之遊，道不奉　　S
　<明>　<德>害<曷>，城<**誠**>胃<謂>之也。又<有>命自天　　S

4．3．俗

单字	频率	文本	左传	论语	墨子	孟子	荀子	韩非	吕氏
俗	159	5	0	0	14	5	79	15	46

A) 单字用法：上博 1 次；郭店：1 次；超出左、论, 少于墨，荀、韩、吕。
B) 固定搭配: 世 X 26; 習 X 9; 成 X 6; X 主 7; 風 X5; 殊 X5; X 儒 5
　　世俗 26: 墨 5, 孟 2, 《荀子》16，《吕氏春秋》3；竹书：无
　　習俗 9: 《荀子》6，《吕氏春秋》3；竹书：无
　　成俗 6: 《荀子》6; 竹书：无
　　風俗 5: 《荀子》5; 竹书：无
C) 结论: 整体上竹书接近战国早期文献。

逐字索引

則君長勞。故君民者，章好以示**民俗**，慎惡以禦民之淫，則民不惑矣.　　G
物>也，尃<溥>　<觀>**人**谷<**俗**>安<焉>，大　<歛>材安<焉　　S

4．4． 擅

单字	频率	文本	左传	论语	墨子	孟子	荀子	韩非	吕氏
擅	70	4	3	0	0	0	20	43	4

A) 单字用法：上博 0 次；郭店：0 次；同左、论、墨、孟。
B) 固定搭配: X 為 6; X 行 5; 得 X 5; X 作 5
　　擅為 6：《韩非子》4，《吕氏春秋》2；竹书：无
　　擅行 5：《韩非子》 4，《吕氏春秋》1；竹书：无
　　得擅（不/莫得擅 V）5：《荀子》1, 《韩非子》 2，《吕氏春秋》2；竹书：无
　　擅作 5：《左传》 1, 《荀子》3, 《韩非子》 1; 竹书：无
C) 结论: 整体上竹书接近战国早期和中期文献。

4．5． 誹

单字	频率	文本	左传	论语	墨子	孟子	荀子	韩非	吕氏
誹	30	3	0	0	0	0	3	9	18

A) 单字用法：上博 0 次；郭店：0 次；同左、论、墨、孟。
B) 固定搭配: X 謗 8
　　誹謗 8：《韩非子》5，《吕氏春秋》3；竹书：无

C) 结论: 整体上竹书接近战国早期和中期文献。

五、小结

词汇分布一致性抽样调查的结果系统地支持了楚竹书语言接近中期或更早文献的结论。当然这完全可以看作是一种事后结论，因为学术界早已通过独立的考古证据把楚竹书大致判定为这个时期的文献了。不过我们的方法的特点是没有参入任何其他的判断标准，完全是以语言用法为参考项，应该有一定程度的科学客观性。

六、应用

为了验证上述方法的有效性，我们尝试把得到的结果运用到其他一些古书的断代上去。我们考察的是尤锐考察过的《孙子》和《商君书》，外加上《国语》、《尉缭子》和《六韬》。所有这些著述都有时代不确定的问题。下面先看一些跟这

些古书有关的研究文献上的结论（重点为我们所加）。有关《孙子兵法》的断代，金精认为：

> 迄今为止,《《孙子兵法》的成书年代尚无定论。根据历史事件的资料考证,推断出《《孙子兵法》十三篇的成书年代应是**公元前 514 年**。[10]

而尹振环指出：

> 本文对何炳棣先生提出的《孙子兵法》早于《老子》之说,进行具体考释,比较《孙》、《老》之思想相通处,进一步证实《孙》早于《老》的说法,并论述《老》从《孙》中借鉴的内容。[11]

关于《尉缭子》一书，徐勇指出:

> 今本《尉缭子》过去也被许多人视为伪书,不敢征引。现在这部兵书经考古发现和研究,已被公认为**战国时期**的作品。[12]

关于《六韬》，有研究者指出：

> 《六韬》：从避讳情况来看,竹简书写的时间应在西汉前期。又所出竹书的字体属于早期隶书,据推断是文、景至武帝初这段时期抄写成的。这一地下出土文献有力证明了此书在西汉以前就广为流传,其成书年代必在**先秦战国之时**。[13]

另有研究者指出：

> 一些学者断定《六韬》的成书时间在战国时期。又有学者根据《史记·留侯世家》的分析，断《六韬》为“秦始皇在位时写成。”还有学者认为《六韬》是**齐桓公时根据周室档案“修复太公法”而成**。我认为，后者是正确的，还认为到齐威王时，齐大夫们再一次对《六韬》进行了修订，从而又杂糅进一些战国时代的特色。[14]

---

[10]金精：《〈孙子兵法〉十三篇的成书年代是公元前 514 年》《西安交通大学学报》25.2（2005）: 96。

[11] 尹振环：《〈老子〉从〈孙子兵法〉中借鉴了什么—也谈〈孙子兵法〉早于〈老子〉》，《学术月刊》, 2004 年 11 期，页 80-86。

[12] 徐勇：《〈吴子〉的成书、著录及其军事思想》，《军事历史研究》2001 年第 3 期, 页 142-148。

[13] 解文超,崔宏艳：《〈六韬〉真伪考》，《青海师范大学学报(哲学社会科学版)》 2005 年第 2 期, 页 73-77。

[14] 仝晰纲：《〈六韬〉的成书及其思想蕴含》，《学术月刊》2007 年第 7 期，页 94-98。

最后，对《商君书》一书，张林祥总结道：

> 关于《商君书》的编成时间、编者及命名等问题一直存在争议。**本文认为它的成书上限应为前260年,而下限则难以断定**;编者当为商鞅后学或商鞅学派无疑,后又经刘向手校,编定为二十九篇。它最初应无名,至刘向校定,始以《商君》为其书名,后又变为《商君书》、《商子》。[15]

上述著作除了把《孙子兵法》明确认定为成书于公元前514年、《商君书》的成书上限明确认为前260年之外,对其他诸书的年代都语焉不详。而尤锐却认为《孙子》比较接近《墨子》和郭店楚简，应属于前四世纪中期左右的作品，而《商君书》则更接近《孟子》和郭店楚简，为公元前四世纪末作品。

如果把前面考察过的先秦文献的几个关键词语的用法跟在这些古书中的用法加以比较，我们会得到一些更清楚的认识。限于篇幅，我们这里还是把考察范围限于巧、誠、俗、擅、誹这几个关键字及其组合格式。

首先，从语词分布的范围看（表乙），《孙子》、《国语》、《尉缭子》早于《六韬》和《商君书》。《孙子》只有“巧”字出现两次，其他都没有出现。其余的作品每个字的用率都大大超出《孙子》。《六韬》和《商君书》则几乎每个字都有用例，因此可以推断它们的晚出性（至少可以说它含有较多的晚出成分）。

其次，从组合格式的用法看，《六韬》和《商君书》的组合式的类型大大超出《孙子》、《国语》、《尉缭子》，而后面三个文本很少见到组合格式。

表甲：先秦7部作品若干单字用法

单字	左传	论语	墨子	孟子	荀子	韩非	吕氏
巧	2	5	13	4	20	37	35
誠	1	2	0	21	66	34	37
俗	0	0	14	5	79	15	46
擅	3	0	0	0	20	43	4
誹	0	0	0	0	3	9	18

[15] 张林祥：《〈商君书〉的成书与命名考辨》，《古籍整理研究学刊》2007年第2期，页1-3。

表乙：先秦另 5 部作品若干单字用法

单字	孙子	国语	尉缭子	六韬	商君书
巧	2	5	0	4	19
誠	0	3	8	7	3
俗	0	0	1	5	16
擅	0	6	0	1	5
誹	0	0	0	1	0

具体组合格式及出现频率如下：

《孙子》：巧， 2；誠，0；俗，0；擅，0；誹，0

《国语》：

巧， 5，巧文 1
誠，3；
俗，0；
擅，6；擅進退 2，擅 舍/捨 2
誹，0

《尉缭子》：

巧， 0
誠，8；不诚 3，精诚 1
俗，1；
擅，0
誹，0

《六韬》：

巧， 4，技巧 1，巧者 1，巧手 1，
誠，7；诚信 2，诚者 2
俗，5；世俗 3，时俗 1，民俗 1
擅，1；擅天下之利者
誹，1：誹時俗

《商君书》：巧，19，技巧 4，巧言 2，世巧 2，知巧 2

誠，3（诚用、诚信、诚任）；
俗，16；成俗 3，俗成 3，世俗 3，异俗 2，时俗 1
擅，5；擅徙 2，擅发 1
誹，0

第三，把上述单字和组合的结果综合起来并与《左传》等文献比较，我们似乎可以推断《孙子》、《国语》、《尉缭子》接近《左传》、《论语》等春秋末期或战国

早期（即公元前 5 世纪或更早）作品。《六韬》则相当于战国前期（公元前 5 世纪末至 4 世纪）的《墨子》（这个结论与仝晰纲的结论较接近）。最后，我们认为《商君书》应该和战国中期（公元前 4 世纪末）的《孟子》和楚简相对。换句话说，我们的结论支持上述著作把《孙子兵法》明确认定为成书于公元前六世纪，不支持把《商君书》的成书上限确定为前三世纪（应该稍早）。我们的结论也不同于尤锐对《孙子》的判断。我们认为《孙子》应该早于《墨子》和楚简，应属于前五世纪左右的作品，但是我们的结论支持尤锐关于《商君书》更接近《孟子》和郭店楚简，为公元前四世纪末作品的看法。也就是说，相关的典籍应该如下断代：

春秋末期或战国早期（前 5 世纪或更早）：《左传》、《论语》、《孙子》、《国语》、《尉缭子》;

战国前期（公元前 5 世纪末至 4 世纪）：《墨子》、《六韬》;

战国中期（公元前 4 世纪末）：《商君书》、《孟子》、楚简;

战国晚期（公元前 3 世纪末或更晚）：《荀子》、《韩非子》、《吕氏春秋》.

七、结束语

通过几个方面的考察，我们得出如下一些初步的结论。首先，一致性分析方法加上逐字索引和搭配研究等方法可以把语词在不同文本中的分布精确地显示出来，为我们进行其他层次的研究提供有用的材料。这种手段完全可以扩大到其他文本研究工作中去，特别是涉及多种文本比较的工作。第二，语料库分析的结果显示，具有关键字和关键词语价值的成分可以是大范围和多类型的，不应该限于个别语词，更不限于名词性成分的范围。例如我们这里核准的关键语词就有动词性的、形容词性的等等。最后，以统计为基础的语言用法作为文本比较的参考项，应该有一定程度的科学客观性，可以拿来作为一种特殊的佐证，帮助我们确古籍的时代特征以及其他类似的用途上。当然我们也不能夸大语言统计的作用，因为能够统计的对象毕竟不是无限的，因此我们必须力争结合多种类型的证据实现对文献的更为科学的判断。

这项研究也有若干明显的缺陷。例如，单字的统计固然有用，但每个单字不一定只有一个意义。在没有区别意义的情况下所作的单字的统计就难免失之偏颇。第二，古书中相互引用的情况相当普遍，不同时代的文本在语言文字特征上不免有重叠相似的地方。如何剥开文本的层次使统计变得更有意义恐怕还得结合其它的手段来进行，比如我们设想，给古书文本中的引语加标，分别统计原书和引语的语言用例可能是今后应该努力的一个方向。[16]

[16] 本段内容吸收了参加香港中文大学 2008 年「《老子》文本两千年之演变及相关简帛资料研究」研讨会诸位先生的批评意见，在此谨致谢意。

附录：(见另页)

附录 1：在七个文本中出现的字项（节选前 30 个）

字项	频率	文本域	左传	论语	墨子	孟子	荀子	韩非	吕氏
便	98	7	1	4	11	1	26	24	31
蹈	16	7	1	2	1	1	5	4	2
恭	88	7	1	13	8	15	32	12	7
徑	26	7	1	1	3	1	5	8	7
口	151	7	1	2	20	16	46	30	36
論	224	7	1	2	7	2	57	68	87
畔	25	7	1	4	2	6	5	1	6
攘	17	7	1	1	5	3	1	2	4
秉	11	7	2	1	1	2	3	1	1
雕	30	7	2	3	1	2	11	8	3
餓	47	7	2	1	5	7	1	25	6
附	21	7	2	1	1	5	8	3	1
槨	20	7	2	3	3	3	4	2	3
接	78	7	2	1	9	7	24	7	28
潔	46	7	2	3	13	3	1	16	8
謹	96	7	2	3	9	2	45	16	19
均	47	7	2	2	9	1	16	7	10
鳴	65	7	2	2	19	5	8	9	20
巧	116	7	2	5	13	4	20	37	35
仞	21	7	2	1	2	1	6	5	4
豚	20	7	2	1	6	4	3	2	2
譽	177	7	2	2	63	3	19	74	14
飽	31	7	3	3	5	8	6	4	2
尺	58	7	3	1	5	12	10	21	6
臭	44	7	3	1	3	1	8	9	19
讀	22	7	3	1	5	1	4	5	3
割	41	7	3	2	2	2	4	19	9
拱	12	7	3	1	1	1	2	3	1

附录 2：在六个文本中出现的字项（节选前 30 个）

字项	频率	文本域	左传	论语	墨子	孟子	荀子	韩非	吕氏
創	9	6	0	1	1	1	1	4	1
飯	33	6	0	6	4	2	3	11	7
膚	26	6	0	3	3	7	4	2	7
稼	47	6	0	3	15	2	4	1	22
矩	41	6	0	1	9	6	7	13	5
窺	23	6	0	1	2	1	2	6	11
倫	31	6	0	2	3	9	10	1	6
沐	14	6	0	1	3	1	3	4	2
版	8	6	1	1	0	1	1	3	1
鼻	32	6	1	0	2	2	10	6	11
避	86	6	1	0	17	4	11	31	22
菜	12	6	1	1	0	1	3	1	5
參	85	6	1	3	3	0	20	47	11
誠	161	6	1	2	0	21	66	34	37
第	124	6	1	20	38	0	32	16	17
兒	28	6	1	1	2	0	1	11	12
拂	29	6	1	0	4	2	10	10	2
槁	16	6	1	0	2	3	5	1	4
供	19	6	1	0	1	6	1	4	6
機	21	6	1	0	4	1	3	8	4
寄	15	6	1	1	2	0	1	8	2
蠡	13	6	1	0	1	1	2	2	6
裸	11	6	1	0	1	2	1	1	5
米	25	6	1	0	2	3	3	2	14
慕	41	6	1	0	3	12	11	12	2
撓	24	6	1	0	1	1	3	13	5
破	52	6	1	0	3	1	3	34	10
殊	29	6	1	0	1	2	1	1	23

附录 3：在五个文本中出现的字项（节选前 30 个）

字项	频率	文本域	左传	论语	墨子	孟子	荀子	韩非	吕氏
俗	159	5	0	0	14	5	79	15	46
錯	54	5	0	4	4	0	30	9	7
勢	251	5	0	0	1	5	27	175	43
肖	173	5	0	0	18	5	26	45	79
怪	41	5	0	1	6	0	17	10	7
全	95	5	0	0	7	1	17	23	47
喻	25	5	0	2	0	3	14	1	5
巨	44	5	0	0	5	5	12	3	19
繆	56	5	0	0	2	6	11	4	33
械	31	5	0	0	3	2	11	3	12
輟	31	5	0	1	2	0	9	12	7
源	16	5	0	0	2	2	9	1	2
智	472	5	0	0	46	32	9	240	145
琢	22	5	0	1	7	2	9	3	0
導	27	5	0	0	3	2	8	1	13
境	79	5	0	0	12	3	8	34	22
符	25	5	0	0	2	1	6	9	7
毫	22	5	0	0	2	2	6	10	2
驥	33	5	0	1	4	0	6	7	15
暖	23	5	0	0	6	2	6	2	7
枉	36	5	0	7	0	6	6	7	10
總	22	5	0	1	6	0	6	1	8
耘	12	5	0	0	1	1	5	3	2
拙	36	5	0	0	1	2	5	17	11
挫	10	5	0	0	1	1	4	3	1
圓	38	5	0	0	1	1	4	11	21
措	15	5	0	1	4	0	3	5	2
暮	28	5	0	0	4	2	3	14	5

附录 4：在四个文本中出现的字项（节选前 30 个）

字项	频率	文本域	左传	论语	墨子	孟子	荀子	韩非	吕氏
曷	46	4	1	0	0	1	28	0	16
調	62	4	0	0	9	0	26	6	21
統	27	4	0	0	0	1	23	2	1
擅	70	4	3	0	0	0	20	43	4
雜	36	4	0	0	5	0	19	7	5
俞	25	4	5	0	2	0	15	0	3
漸	22	4	1	0	1	0	14	6	0
妖	40	4	10	0	2	0	14	0	14
誕	16	4	1	0	0	0	13	1	1
裕	18	4	0	0	0	1	12	2	3
妒	22	4	1	0	0	0	11	9	1
僻	28	4	0	0	3	0	11	8	6
蹠	38	4	0	0	0	4	11	16	7
析	51	4	25	1	0	0	10	0	15
壙	12	4	1	0	1	1	9	0	0
恣	32	4	0	0	0	1	9	8	14
篡	19	4	0	0	8	2	8	0	1
砥	16	4	0	0	2	0	8	5	1
睹	22	4	0	0	8	1	8	0	5
樸	28	4	0	0	4	0	8	10	6
笙	16	4	1	0	6	0	8	0	1
竽	27	4	0	0	7	0	8	9	3
粥	12	4	1	0	0	2	8	0	1
絀	12	4	1	0	0	0	7	1	3
黻	11	4	2	1	0	0	7	0	1
穢	16	4	4	0	0	0	7	3	2
漏	21	4	2	0	4	0	7	8	0
蔔	126	4	96	0	0	0	6	10	14

附录 5：在三个文本中出现的字项（节选前 30 个）

字项	频率	文本域	左传	论语	墨子	孟子	荀子	韩非	吕氏
綦	41	3	1	0	0	0	38	2	0
案	39	3	0	0	0	0	35	3	1
愨	33	3	0	0	0	0	29	3	1
愉	29	3	0	2	0	0	19	0	8
詘	33	3	0	0	0	0	17	6	10
篇	23	3	0	0	4	0	17	0	2
悍	20	3	0	0	0	0	13	4	3
踰	15	3	1	0	0	2	12	0	0
技	21	3	1	0	0	0	10	10	0
按	19	3	0	0	0	0	10	2	7
俄	18	3	0	0	0	0	10	7	1
粹	13	3	0	0	0	0	10	1	2
態	13	3	0	0	0	0	10	1	2
恬	17	3	0	0	1	0	9	7	0
黼	11	3	1	0	0	0	9	0	1
影	17	3	0	0	0	0	8	1	8
漫	14	3	0	0	0	0	8	1	5
贈	22	3	9	0	0	0	7	6	0
枯	16	3	0	0	0	0	7	2	7
誇	14	3	0	0	0	0	7	4	3
芬	11	3	0	0	1	0	7	0	3
驗	28	3	0	0	0	0	6	10	12
蘭	20	3	11	0	0	0	6	3	0
唱	18	3	0	0	0	0	6	4	8
歟	18	3	0	0	0	1	5	0	12
竄	12	3	5	0	0	0	5	0	2
忒	11	3	0	0	0	0	5	1	5
苴	9	3	3	0	0	0	5	0	1

附录 6：在两个文本中出现的字项（节选前 30 个）

字项	频率	文本域	左传	论语	墨子	孟子	荀子	韩非	吕氏
邾	132	2	126	0	0	0	0	0	6
僖	74	2	68	0	0	0	0	6	0
戊	66	2	64	0	0	0	0	0	2
巳	62	2	43	0	19	0	0	0	0
囗	61	2	17	0	0	0	44	0	0
僑	41	2	38	0	0	0	0	3	0
埶	40	2	0	0	0	0	39	0	1
郤	39	2	0	0	0	0	0	19	20
俘	38	2	37	0	0	0	0	0	1
賄	37	2	35	0	0	0	0	0	2
酉	31	2	30	0	0	0	1	0	0
樊	29	2	18	11	0	0	0	0	0
印	26	2	0	0	0	0	5	0	21
宓	25	2	0	0	0	0	0	3	22
葆	24	2	0	0	12	0	0	0	12
郎	23	2	8	0	0	0	0	15	0
鄴	23	2	0	0	0	0	0	13	10
這	23	2	0	0	0	0	1	22	0
僭	22	2	17	0	0	0	5	0	0
哪	22	2	21	0	0	0	0	1	0
它	22	2	0	0	0	0	20	0	2
刖	22	2	6	0	0	0	0	16	0
竅	21	2	0	0	0	0	0	7	14
閽	20	2	0	0	0	0	18	0	2
褚	20	2	18	0	0	0	0	2	0
刁	19	2	0	0	1	0	0	18	0
攜	19	2	17	0	0	0	0	0	2
只	19	2	16	0	3	0	0	0	0

附录 7：在一个文本中出现的字项（节选前 30 个）

字项	频率	文本域	左传	论语	墨子	孟子	荀子	韩非	吕氏
戌	86	1	86	0	0	0	0	0	0
丐	33	1	33	0	0	0	0	0	0
蒯	29	1	29	0	0	0	0	0	0
紇	23	1	23	0	0	0	0	0	0
乜	23	1	23	0	0	0	0	0	0
榿	22	1	22	0	0	0	0	0	0
燮	21	1	21	0	0	0	0	0	0
縊	21	1	21	0	0	0	0	0	0
櫟	15	1	15	0	0	0	0	0	0
鄆	14	1	14	0	0	0	0	0	0
俾	13	1	13	0	0	0	0	0	0
魴	13	1	13	0	0	0	0	0	0
麋	13	1	13	0	0	0	0	0	0
郯	13	1	13	0	0	0	0	0	0
頹	13	1	13	0	0	0	0	0	0
媯	12	1	12	0	0	0	0	0	0
閏	12	1	12	0	0	0	0	0	0
莞	11	1	11	0	0	0	0	0	0
硌	11	1	11	0	0	0	0	0	0
裔	11	1	11	0	0	0	0	0	0
聹	10	1	10	0	0	0	0	0	0
甓	10	1	10	0	0	0	0	0	0
腆	10	1	10	0	0	0	0	0	0
舴	10	1	10	0	0	0	0	0	0
娣	9	1	9	0	0	0	0	0	0
諜	9	1	9	0	0	0	0	0	0
饃	9	1	9	0	0	0	0	0	0
翩	9	1	9	0	0	0	0	0	0