

JOURNAL OF

CHINESE LINGUISTICS

VOLUME 41, NUMBER 1

JANUARY 2013

ARTICLES

- Jingxia Lin:
Factors restricting the use of the locative preposition *yu* in classical Chinese: A modeling of recursive partitioning (in Chinese)..... 1
- Daniel W. C. So and Chun-fat Lau:
Rapid large scale intra-nationality language shift in Hong Kong 1949-1971..... 21
- Kawai Chui:
Gesture and embodiment in Chinese discourse..... 52
- Yicheng Wu:
On the so-called unbounded passives..... 65
- Huili Zhang and Haihua Pan:
The predicate inflection in Yancheng dialect (in Chinese)..... 91
- Shuling Huang and Kehjiamn Chen:
Semantic analysis and contextual harmony of durations..... 118
- Bin Li:
Integrating textual and prosodic features in the interpretation of Chinese utterance-final particles: A case of *a* and *ne*..... 145
- Chinfa Lien:
The polyfunctionality of *leio* 叫 in *niri kiam khi* 荔枝記: A frame-net-based approach..... 170
- Quansheng Xia, Yong Lü, Xuejun Bai, Feng Shi:
An ERP study on the semantic processing of unambiguous noun, unambiguous verb and verb-noun ambiguous word in Mandarin Chinese (in Chinese)..... 197

MEMORIAM

- W. South Coblin:
Jerry Norman: Remembering the man and his perspectives on Chinese linguistic history.....

REPORT

.....

Complimentary copy

提要

有关制约古代汉语地点介词“於/于”应用的种种条件，在学术界（方平权（2000）、董秀芳（2006）、Xu（2006）、许敏云（2008）等）有很多不同的说法，至今尚未有定论。本文通过对历史语料进行“递归分区”的分析，对所有现行相关的解释进行了测试，用统计学的手段，去伪存真。本文发现在古代汉语中，“於/于”的使用或省略首先与动词的及物性相关。高及物动词倾向于使用“於/于”来引出地点名词，而低及物动词则不是如此。该结果刚好与董（2006）和 Xu（2006）相悖。汉语历史时期是第二个重要影响因素。随着汉语的历史发展，尤其是在《水浒传》以及之后的文献中，“於/于”使用频率逐步减少。另外一个比较重要的因素是地点名词的音节数，单音节地点名词倾向于用“於/于”引出。但是，用强调、抽象地点名词、韵律等等说法来解释“於/于”的相关用法，统计测试结果表明可能是没有科学依据的。本文提供的数据为进一步研究“於/于”在历史文献中的语法功能，建立了一个具有科学依据的基础。¹

关键词： 地点介词 “於/于” 递归分区统计分析法

1. 背景

本文使用统计方法评估制约古代汉语中地点介词“於/于”使用的所有相关因素。²关于“於/于”的使用和省略这一问题已经有多位学者提出过解释，比如方平权（2000）、董秀芳（2006）、Xu（2006），以及许敏云（2008）等等，但是学术界至今尚未有定论，甚至还存在截然相反的说法。而对于为什么学术界会在诸多不同的解释，也是一个未知问题。因此，

本文采用递归分区方法分析历史数据,从统计角度衡量各种可能解释,去伪存真。本文发现在古代汉语中,“於/于”的使用或省略首先与动词的及物性相关。高及物动词倾向于使用“於/于”来引出地点名词,而低及物动词则不是如此。汉语历史时期是第二个重要影响因素。随着汉语的历史发展,尤其是在《水滸全传》以及之后的文献中,“於/于”使用频率逐步减少。另外一个比较重要的因素是地点名词的音节数,单音节地点名词倾向于使用“於/于”。之名词化,隐喻,强调等因素在统计上的重要性相对比较小。本文提供的数据为进一步研究地点介词“於/于”在历史文献中的语法功能,建立了一个具有科学依据的基础。

“於/于”被其他介词比如“在”取代之前,我们能在古代汉语中找到不少使用或省略“於/于”的例子。(1a)出自《史记》,“於/于”被用来引出地点名词“吴中”,而(1b)出自《汉书》,虽然描述的事件和(1a)基本一致,但是省略了“於/于”。

- (1) a. 项梁杀人,与籍避仇於吴中 (史记)
b. 梁尝杀人,与籍避仇吴中 (汉书)

不使用“於/于”的地点结构和使用“於/于”的地点结构在功能上基本一致。两者都能引出地点名词来表示非方向性地点,如(1a-b);目的地,如(2a-b);来源,如(3a-b),但是表目的和来源的例子要明显少于非方向性地点。

- (2) a. 他赶潮头,来於此处 (西游记)
b. 吾所以还军霸上 (史记)
(3) a. 君取於吴,为同姓,谓之吴孟子 (论语)
b. 征兵九江王布 (史记)

关于“於/于”的使用或省略,学术界已经提出多种解释。如吴国忠(1989)认为“於/于”的使用具有随意性;董秀芳(2006)认为不使用“於/于”能强调后面的地点名词;而许敏云(2008)认为使用“於/于”才能强调后面的地点名词。这样一来,问题就产生了:如果所有的说法都是合理的,那么各种说法之间存在什么样的关系,尤其是像董秀芳(2006)和许敏云(2008)这样相反的说法?如果连两个相反的说法都是合理的,那么各种说法之间会不会存在一种层次体系,从而能优化“於/于”的使用?

为了回答这些问题,我们需要找出哪个因素最能决定“於/于”的使用。不管是从语义角度还是句法角度,前人提出的每个说法都有一定的实例支持,但是也都存在一些例外数据。从统计学的角度来说,少量例外(“outliers”)的存在并不奇怪,单独一个因素也不能解释所有数据,但是我们可以使用合适的统计模型,找出能解释最多数据的因素。本文中,我们采用一个叫做“递归分区”(“recursive partitioning”,也叫“classification and regression trees”)的数据挖掘算法,来评估所有因素在统计上预测“於/于”使用或省略的能力,并找出能在最大程度上解释数据的那个因素。

下文中,第二部分描述前人提出的所有相关说法;第三部分介绍统计分析所需的数据和方法,除了学术界已存在的说法,本文也会包括另外三种可能的解释,这三种解释是作者在数据收集过程中发现的;第四部分分析并讨论统计得出的结果;第五部分是全文总结。

2. 前人研究

关于“於/于”的使用或省略研究,相比较“於/于”的功能及历史发展研究而言,数量比较少,但是研究角度多种多样。现有的研究文献可以大致分为五类。

2.1 语言经济和简化原则

第一类说法认为“於/于”的省略是语言经济简化的结果。何乐士(1985)针对《左传》和《史记》中的简体“于”和繁体“於”做过一次数量统计。如表1所示,从《左传》到《史记》,两种形式在使用频率上都有所下降:“於”从7.8%到5.7%，“于”从7.3%到0.1%。根据何乐士的这一结果,方平权(2000)指出语法除了可以逐渐缜密化之外,还可以不断简化。“於/于”作为使用频繁的介词,在汉代文献中的省略应该是其从先秦发展而来。

表1 《左传》和《史记》中的“於/于”在动词,形容词后的使用频率(何乐士1985)

	《左传》		《史记》	
	频率	千分比	频率	千分比
於	1534	7.8	417	5.7
于	1442	7.3	8	0.1

时兵 (2003) 也提出过一个类似的说法。他认为“动词 + 於/于 + 与事/位事名词”是原型双宾结构的早期语法格式, 而“动词 + 与事/位事名词”是由于较高的使用频率和语言经济原则的作用, “於/于”逐渐脱落而来。

尽管如此, 我们可以发现方平权和时兵都只考虑了“於/于”在一定历史时段衰落的现象。在现代汉语中, 我们可以看到地点介词或虚词并没有变得更加经济简化。相反, 新的介词产生并代替了“於/于”, 且这些新介词是不可省略简化的。如在 (4) 中, 如果非方向性介词“在”被省略, 句子就不能成立。若语言经济简化原则是合理的, 那么在古代汉语中存在的介词将会逐渐衰落, 而新的介词也不会产生。

- (4) a. 我在学校看书
b. *我学校看书

2.2 写作风格

第二类说法由吴国忠 (1989) 提出。他认为在古汉语中, “於/于”的使用存在不同程度的随意性, 且该随意性不是由不同时期不同人的不同语言使用习惯导致的, 因为同一部作品中也能发现“於/于”的随意性。此外, 吴国忠也指出古汉语时期, 人们没有严格的语法观念, 因此对这样的随意性不难理解。但是, 吴国忠的说法难以解释如上面表 1 中的数据, 即为什么在先秦文献如《左传》中, “於/于”的使用频率高于汉代文献如《史记》, 也不能解释为什么“於/于”的使用在晚期上古汉语和中古汉语中衰退后, 又在近代汉语中变得频繁起来。

2.3 之名词化和音节数

第三类说法涉及地点名词的名词化和音节数。Xu (2006) 观察到, 当动词后存在复杂名词 (complicated) 时, “於/于”倾向出现。根据 Xu (2006) 的定义, 复杂名词指由短语被“之”名词化后所成的名词性成分, 如 (5a); 相对地, 若地点名词是简单名词 (simple), 如单音节名词, “於/于”则可用可不用, 如 (5b)。

- (5) a. 秦王之怒於楚之缓也 (战国纵横家书) (Xu 2006, 86)
b. 楚王怒周 (战国策) (Xu 2006, 87)

然而 Xu (2006) 的这一说法也不能系统地解释所有数据, 因为也存在大量的反例。(6) 是其中一个反例: 动词“怒”后面的名词“公子之盗其兵符”, 不但是多音节名词, 而且是由“之”名词化而来, 但是“於/于”还是没有出现。

- (6) 魏王怒公子之盗其兵符 (史记) (Xu 2006, 89)

2.4 及物性变化

第四类说法和动词的及物性有关。许敏云 (2008) 指出, 在汉代, 如《史记》时期, 像“献”和“告”这样的动词已经获得了带间接宾语的能力, 所以不再需要使用“於/于”来引出间接宾语。Xu (2006) 也讨论过及物性对“於/于”使用的影响。Xu (2006) 指出不使用“於/于”的时候, 原来跟在“於/于”后面的名词变成动词的宾语。根据许敏云 (2008) 和 Xu (2006) 的说法, 我们可以推出若不使用“於/于”, 原来不及物的动词会变成及物动词, 而原来单及物的动词会变成双及物动词。因此, (7a) 中“魏”是“於”地点介词的宾语, 而 (7b) 中的“云阳宫”则是动词“死”的宾语。(7c) 是 (7b) 在现代汉语中的表达方式, 但是我们可以看到不管地点介词“在”出现在动词前还是动词后, “云阳宫”都是“在”的宾语。

- (7) a. 张伊死於魏 (史记)
b. 夫人死云阳宫 (史记)
c. 夫人在云阳宫死去/夫人死在云阳宫

2.5 强调

第五类说法是从语用学角度提出的。除及物性外, 许敏云 (2008) 也指出“於/于”的使用对其后的名词性成分有明显的强调作用。比如, 许敏云认为 (8a) 着重描述事件发展的动态过程, 并突出彭城的交与对象“宋”, 而 (8b) 着重描述事件发展的静态过程, 突出对彭城的获得。

- (8) a. 归彭城于宋 (Xu 2008, 90)
b. 归我彭城 (Xu 2008, 90)

同样从语用学角度, 董秀芳 (2006) 的观点和许敏云 (2008) 在一定程度上互相对立。根据 Hopper & Thompson (1980), 董秀芳指出低及物性动词倾向用“於/于”引出名词, 而高及物动词则直接带名词。至于那些低及物

动词直接带名词的反例，董秀芳认为在这些例子中，“於/于”是被有意地省略掉，从而达到强调名词的目的，且这些名词也变成了动词的临时宾语。

然而，不管是因为“於/于”还是省略，被“强调”的名词都处于短语末尾位置 (cf. Lambrecht 1994)。也就是说，即使“於/于”被省略了，这个名词也没有发生真正的位移。因此，不管“於/于”使用还是不使用，都没有对其后的名词起强调性作用。

如上所述，尽管学者们已尝试从不同角度来提出制约“於/于”在古代汉语中使用的因素，但是没有一个是能系统地解释所有例子。因此，我们需要采用合适的方法来评估这些因素。由于这些因素是从不同角度提出来的，一个理想的评估方法应该量化地衡量所有因素，并找出在统计上最有意义的因素。对历史数据进行“递归分区”分析是一个能满足这样需求的统计方法。

3. 研究方法

3.1 数据

如表 2 所示，在 Sun (2006) 关于汉语历史时期划分的基础上，本文将先秦时期到清代的汉语历史分为七个时期，并在每个时期选取字数相当的文献来查看“於/于”的用法。鉴于口语在语言变化中的重要性，本文在选取时偏重较接近口语的文献。本文一共选取了九部文献，总字数约为 258.5 万字。

表 2 考察文献

文献	汉语时期划分	历史时期	字数
《左传》	早期上古汉语	先秦 403 BCE前	约 27.5 万
《史记》	晚期上古汉语	汉 206BCE-220CE	约 50 万
《世说新语》	早期中古汉语	南北朝 420 - 589	约 8 万
《六度集经》			约 6 万
《贤愚经》			约 16 万
《敦煌变文》	晚期中古汉语	唐 618-907	约 23 万
《水浒传全传 (41-80 回)》	早期近代汉语	明初 1368-1644	约 33 万

《喻世明言》 ³	早期近代汉语	晚明 1368-1644	约 36 万
《女儿英雄传》	晚期近代汉语	清 1644 to 1911	约 59 万

该九部文献中，所有符合“动词 + 宾语/Ø + 於/于/Ø + 地点名词”结构的例子全部被搜索并收集起来做统计分析。⁴ 尽管“於/于”带地点名词可以修饰大量不同类型的动词，本文从可操作性角度考虑，仅选择考察了六个动词：死，怒，杀，取，教，遗。需要注意的是该六个动词能反映及物性以及语义的多样性。本文对及物性的理解是尽管表现形式不同，一个动词内在的及物性是固定的。例如“死”表达死亡的动作时不会受影响涉及主语之外的人或物，因此“死”从内在角度来说是不及物动词；“教”表达教授的动作时必然涉及三方面：教的主语，被教的人以及教的内容，因此“教”从内在角度来说也是双及物动词。本文把及物性分为低及物，高级物和双及物，其中低级物包括不及物动词和心理动词等及物性较弱的动词。针对每类及物，本文选择考察两个动词：“死”和“怒”作为低及物动词，“杀”和“取”作为高及物动词，“教”和“遗”作为双及物动词。⁵ 在语义上，这些动词也能反映普通动作（例如给与类动词）和心理动作（例如心理类动词）。⁶ 此外，该六个动词都是常见动词，在现代汉语之前出现频率较高，因此不会有由于语言历史变化而产生的较大的频率差异。另外一个重要原因就是这些动词都能在必要的时候带地点介词短语。表 3 列出了该六个动词的情况。

表 3 考察动词

动词	及物性	语义类型	介词短语中的名词类型
死	低	死亡类 (dying)	物理地点 (Physical location)
怒	低	心理类 (psychological)	刺激物 (Stimulus)
杀	高	谋杀类 (murder)	物理地点 (Physical location)
取	高	获取类 (obtaining)	提供者 (Provider)
教	双	教授类 (teaching)	接受者 (Recipient)
遗	双	给予类 (giving)	接受者 (Recipient)

搜索“死/怒/杀/取/教/遣 + 宾语/Ø + 於/于/Ø + 地点名词”结构的工作完成后，一共收集到 293 条例子。其中在《儿女英雄传》中，没有任何符合该结构的例子出现，但是有大量“动词+宾语/Ø + 在 + 地点名词”的例子。该现象表明在现代汉语中，新的介词“在”可能已代替“於/于”。此外，我们也没有“动词 + 地点名词”的例子，这说明“在”在现代汉语中是不可省略的。

3.2 递归分区统计分析

递归分区是一种数据挖掘算法，其目的是将因变量 Y(dependent variable) 和一系列独立变量 X(independent variable，也被称为“因素”(factors, predictors)) 联系起来，从而得出两者之间的概率关系 $P(Y|X)$ ，即在已知 X 的情况下，Y 事件发生的概率是多少 (Zhang & Singer 1999, Baayen 2008)。

递归分区进程将原始数据以递归形式划分为组，直到进程中找不到在统计上有意义的组为止。在每一次递归分区中，所有独立变量 X 都由算法检测一遍，从而找出能在最大程度上把数据分区的因素，并将这个因素作为分区的标准。决定分区的统计标准是 p-值 (5%)。⁷

递归分区的典型结果是树形图。该树形图从包含全部数据的根节点出发，若有某一个因素能将数据从统计最大化角度分区，则该根节点被该因素分区为两个子节点，每个子节点包含各自的数据，这就是第一次递归分区。第一次分区后的两个子节点将各自进行第二次对所有因素的计算，看某一个因素是否能再次在统计上将子节点分区，若是能再分区，则该子节点再次被分区为两个子节点。这样的递归算法一直持续到没有任何因素能将子节点中的数据划分到两个有意义的分区中为止。

对本文“於/于”的数据以及制约因素来说，递归分区是一个理想的统计方法。首先，它同时允许 X 和 Y 的数据类型为分类数据和顺序数据。⁸ 第二，用递归分区的方法不但能从 X 因素中预测到 Y 发生的概率，还能从统计的角度得出和 Y 关联最大的那几个因素。因此，如果每个因素都和“於/于”的数据有或多或少的关联，那么递归分区会找出能解释最多数据的那几个因素。第三，递归分区得出的树形图简单易懂，因此不要求读者们具备严格的统计或数学背景。

3.3 预测“於/于”使用的因素

在本次统计分析中，如果一个动词带“於/于”介词短语，因变量 Y 就标记为“有於/于”；相反，如果不带介词短语，就标记为“无於/于”。“有”和“无”是分类值 (categorical value)。用来预测 Y 的因素有七个，如表 4 所示。七

个因素中，有四个是用来测试第二部分中介绍的前人的说法，另外三个因素是在数据收集过程中发现的，他们对“於/于”的使用可能也有一定的制约作用。下面将详细介绍每个因素。

该七个因素中，“时间”用来测试作者的写作风格(如吴国忠(1989))。如表 4 所示，每个时间值都代表一个或几个作者的文献以及这些作者所处的历史时期。如果吴的说法是正确的，那么统计分析结果将不会显示“时间”是一个重要预测因素。因为如果来自不同历史时期的作者都是根据自己的习惯随意使用或者省略“於/于”，那么整个数据就不会在时间基础上被清楚地细分开来。相反，如果“时间”是一个重要的统计因素，那就意味着在汉语历史上会出现一个“於/于”的使用频率趋向。不同历史时期的数据也能测试方平权(2000)和时兵(2003)关于语言简化和经济的提议。根据他们的说法，汉语在后期会变得更加简洁经济。如果提议是正确的，那么历史后期中，“动词 + 宾语/Ø + 於/于 + 地点名词”的例子在使用频率上会低于“动词 + 宾语/Ø + 地点名词”。

“及物性”考察动词及物性与“於/于”使用的关系，该因素可以测试许敏云(2008)和 Xu(2006)。““於/于”后音节”和“之名词化”考察 Xu(2006)，即“於/于”的使用是否受到地点名词的音节数以及“之”名词化影响。七大因素中不包括董秀芳(2006)和许敏云(2008)所提出的“强调”作用。原因已在上文讨论过，即“於/于”的使用或省略不会真正对其后面的地点名词起到强调作用。

表 4 预测“於/于”使用的因素

Xs	说明	数值类型	数值标记
时间	标志不同汉语历史时期的文献	顺序数据	1 = 《左传》；2 = 《史记》；3 = 《世说新语》《六度集经》《贤愚经》；4 = 《敦煌变文》；5 = 《水浒传》；6 = 《喻世明言》；7 = 《儿女英雄传》
及物	动词的及物性	分类数据	0 = 低及物；1 = 高及物；2 = 双及物
“於/于”后音节	“於/于”后的地点名词是否多音节	分类数据	0 = 否；1 = 是

之名词化	地点名词是否由“之”名词化而来	分类数据	0 = 否; 1 = 是
“於/于”前音节	“於/于”前动词或名词是否多音节	分类数据	0 = 否; 1 = 是
地点隐喻	地点名词是否有隐喻含义	分类数据	0 = 否; 1 = 是
文体修饰	“於/于”的使用或省略是否出于修饰目的, 如押韵, 对称等	分类数据	0 = 否; 1 = 是

表 4 中的其他三个因素““於/于”前音节”, “地点隐喻”, “文体修饰”没有在前人的研究中详细探讨过。但是在收集数据的过程中, 本文作者发现这几个因素可能与“於/于”有一定的联系。““於/于”前音节”指的是介词“於/于”所在位置前动词或名词是否多音节。如“死於木綿巷”中“於”前面的动词是“死”, 因此是单音节, 而“杀幽王驪山下”中介词位置前的名词是“幽王”, 因此是多音节。如上文介绍, Xu (2006) 指出复杂名词倾向于由“於/于”引出。此外, 在研究介词“在”带地点名词的制约因素中, Sun (2008) 发现“在”后面的地点名词只能是多音节地点名词。两者提示本文, 除了介词后地点名词的音节数, 介词前动词或名词的音节数量也可能影响到介词的使用或省略。在查看动词“死”的数据时, 我们也发现一些相反的例子。如 (9a) 中, 动词“死”和“勒”连用, 组成多音节连动词时, “於/于”不出现; 而 (9b) 中, “死”单独使用, “於/于”出现。这个例子虽然和 Xu (2006) 以及 Sun (2008) 相反, 但是我们也不能忽略它可能的影响, 所以本文将““於/于”前音节”视为一个可能因素。

- (9) a. 將紅羅勒死宮中 (喻世明言)
 b. 直到後來, 死於木綿巷, 方應其語 (喻世明言)

此外, 也有一系列例子显示当地点名词包含隐喻之义时, “於/于”经常出现。如 (10) 中, “奸贼之手”和“黄泉”都不是主语死亡的真实地点。地点介词“於/于”的使用也可能是为了给“於/于”后面的非真正地点名词增加地点色彩, 从而方便读者理解。

- (10) a. 强似死于奸贼之手 (喻世明言)
 b. 如今随仁兄同死於黄泉 (水浒传全传)

为了达到韵律上的平衡或对称, 古代汉语也经常在不影响理解的前提下增加或者省略某些词汇。因此“於/于”的使用也可能受到这方面的影响。比如在 (11) 中, 为了让前后两个短句在字数上相等, 第一个短句没有使用“於/于”, 而第二个短句使用“於/于”。

- (11) 安危在是非, 不在於强弱 (韩非子)

综上所述, ““於/于”前音节”, “隐喻”, “文体修饰”在一定程度上可能与“於/于”的使用有联系, 因此它们和其他因素一起, 都被用来进行统计分析。需要注意的是增加这三个因素不会影响其他几个因素的统计结果。如果某个因素没有统计意义, 该因素就不会显示在树形图中。统计标志的例子可参考附录 A。

4. 结果与分析

4.1 双及物动词“遣”和“教”

六个动词中, “遣”和“教”没有带任何“於/于”介词短语。两个动词都出现在双及物结构 (“遣/教 + 间接宾语 + 直接宾语”) 或单及物结构中 (另外一个宾语省略), 如 (12) 中所示。Rappaport Hovav & Levin (2008, 160) 曾指出, 如果一个语言不使用同一个形态介绍接受者 (recipient) 和空间目的地 (spatial goal), 那么该语言中的给予类动词不会带向格 (allative case)。然而在古代汉语中, 尽管接受者和空间目的地是由同一个形态“於/于”介绍的, 给予类动词也还是不带“於/于”。这个现象表明古代汉语里的给予类动词在与格选择 (dative alternation) 上与其他语言存在类型学上的区别, 因而值得更进一步的探究。不过在本文的递归分区统计分析中, “遣”和“教”不被包括在内。

- (12) a. 上乃使人厚遣关氏 (史记)
 b. 子教寡人伐吴七术 (史记)

4.2 递归分区结果

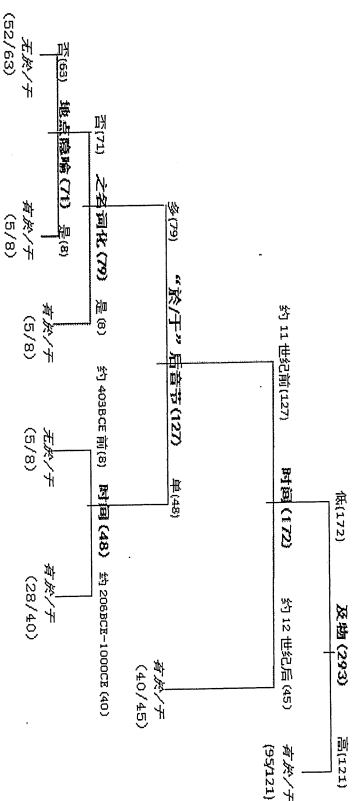
除去“遣”和“教”, 本文对包含剩余四个动词的数据进行统计递归分区。整个分析过程在统计分析软件 R 中完成 (R 是免费软件, 可以在

<http://www.r-project.org/> 下载。附录 B 显示了本次分析在 R 中的编辑语言以及数字结果)。图 1 是分析结果树形图。在该树形图中,根节点从树顶开始,一共有 293 条数据有待递归分区。在树形图中出现的因素有“及物”,“时间”,“於/于”后音节,“之名词化”,“地点隐喻”。这些因素出现在树形图的每个根节点上,表示它们在统计上是为重要的因素。另外两个因素“於/于”前音节和“文体韵律”没有显示,则表示它们不能有效地预测“於/于”的使用。下文将详细讨论该树形图。

如图 1 所示,树形图的第一个节点是“及物”。这表示动词的及物性是影响“於/于”使用的最重要因素。在总共 293 条数据中,如“及物”的右面子节点所示,高及物动词“杀”和“取”的例子有 121 条,其中有 95 条(78.5%)使用了“於/于”。而在左面低及物动词“死”和“怒”的 172 条例子中,有 90 条数据(52.3%)使用了“於/于”。⁹ 董秀芳(2006)指出低及物动词通常需要—个介词来介绍出名词,但是 52.3% 的比例并没有明显显示出低及物动词“於/于”的倾向。董秀芳(2006)和 Xu (2006)也都指出心理动词不对宾语产生直接影响,因此是低及物性动词,倾向带“於/于”。但是通过查看动词“怒”的具体例子,我们也能发现不一样的现象。《史记》一共有 39 条“怒”带名词的例子,但是只有 3 条例子是“怒+於/于+名词”结构,而其他的“怒”都是直接带名词。(13)是其中的两条例子。

- (13) a. 楚,赵怒於魏之先已也 (史记)
b. 鲁桓公知之,怒夫人 (史记)

图 1 树形图——“死”,“怒”,“杀”,“取”数据的递归分区结果



我们已经知道在低及物动词的数据中,使用“於/于”和不使用“於/于”的数据几乎各占一半,这表明“及物”这个因素不能有效预测低及物动词使用“於/于”的情况。因此,树形图中出现了第二次递归分区。在这次分析中,“时间”是分界因素。这一点与吴国忠(1989)的观点相左。如第三部分所讨论的,如果吴的观点是正确的,那么“时间”就不会是一个重要的因素而出现在树形图中。

“时间”将所有 172 条低及物动词的数据按照历史分为两组,右面子节点显示的是约 12 世纪后的文献,即《水浒传传(41-80 回)》,《喻世明言》和《儿女英雄传》,左面子节点显示的是约 11 世纪前的文献,即《敦煌变文》,《贤愚经》,《六度集经》,《世说新语》,《史记》,《春秋左传》。

在右面子节点中共只有 45 条数据。相对于左面子节点的 127 条来说,在 12 世纪后,“动词+宾语/Ø+於/于/Ø+地点名词”的结构使用频率减少。但是在该 45 条数据中,有 40 条(88.9%)使用了“於/于”。这说明尽管总体数量减少,“动词+宾语/Ø+於/于+地点名词”的频率明显高于“动词+宾语/Ø+Ø+地点名词”。我们再回头看原来的数据,就能发现这 40 条带“於/于”的数据全部来自低及物动词“死”,而没有一条来自另外的低及物动词“怒”。其中很大一个原因是“怒”在 12 世纪后在使用方法上发生了变化,不再出现于“怒+於/于/Ø+名词”结构中。

为了证明确实是由于词汇使用上的变化而导致这个结果,我们再看高及物动词“杀”和“取”的数据。“杀”和“取”这两个动词与“死”一样,它们的用法和意义在汉语历史上都没有发生过重大的变化,所以具有可比性。经过查看发现,12 世纪后的“杀”和“取”也基本不出现在“动词+宾语+於/于+地点名词”结构中。¹⁰因此,12 世纪后,动词“死”和其他三个动词的表现很不一样,只有“死”出现在“动词+宾语+於/于+地点名词”结构中。为了证明“死”是一个特殊的动词,我们又考查了另外一个低及物动词“坐”。经过数据统计发现,在《水浒传传(41-80 回)》中,一共只出现三条“坐+於/于+地点名词”的例子,而在《儿女英雄传》中没有“坐+於/于+地点名词”的例子。因此说明在 12 世纪后,“动词+宾语/Ø+於/于+地点名词”结构频率明显降低甚至到无。从另外一方面看,新的地点介词“在”在使用频率上明显提高。“坐+在+地点名词”在《水浒传传(41-80 回)》中,一共出现了 36 条,而在《儿女英雄传》中,出现了 67 条,已经完全代替了“坐+於/于+地点名词”。因此,我们可以得出结论,“於/于”在 12 世纪后的逐步减少不是因为语言经济简化,而是因为新的地点介词“在”的出现并逐步代替了“於/于”。而且“在”作为地点介词在晚期近代汉语中是不可省略的,因为在《儿

女英雄传》中,没有任何“坐+地点名词”的例子被发现。因此,尽管我们暂时不清楚为什么“死”在12世纪后还频繁出现在“死+於/于+地点名词”结构中,其他几个动词表明中古后,“动词+於/于+地点名词”结构的使用频率已开始大量减少,而“动词+地点名词”更是稀少。

“时间”左面子节点的数据还有127条,第三次的统计递归分区发现地点名词的音节数是第三个最为重要的因素。“音节”将数据分成单音节和多音节两组。从右面单音节子节点的末端看,一共有48条数据,其中31条(64.6%)使用“於/于”;而在左边多音节子节点的末端看,一共有79条数据,仅有21条(26.6%)使用了“於/于”。因此,我们可以发现“於/于”可能倾向于出现在单音节地点名词前。这个结果和Sun(2008)关于“在+地点名词”的发现有所不同。Sun(2008)指出现代汉语中的地点介词“在”在一般情况下必须带多音节地点名词,而在古代汉语中,在句法功能和意义上类似“在”的“於/于”却是更倾向于带单音节地点名词。这个有趣的对比现象可能和汉语在历史上的变化有关,但是具体原因还需要进一步的调查研究。同时该结果也表明Xu(2006)关于“之”名词化的说法也是有待商榷的,因为被“之”名词化的名词性成分一定是多音节的。

图1显示“音节”的左右两个子节点又再次进行递归分区。在右面子节点(单音节)中,“时间”再次成为分界因素。但是这次是将10世纪前(从文献角度是《敦煌变文》前)的文献分成两组,左面是《春秋左传》,右面是《史记》,《世说新语》,《六度集经》,《贤愚经》和《敦煌变文》。在右面,一共有40条带单音节地点名词的数据,其中28条(70%)使用“於/于”。这说明“於/于+单音节地点名词”多发生在《春秋左传》之后,《敦煌变文》之前的文献中。查看原始数据后,我们发现这个现象主要出现在《史记》中。如Sun(2008)发现,单音节地名从商代开始大量减少,而由Tan(1982)制作的东汉地图中,已基本没有单音节的地方名。在《世说新语》,《六度集经》,《贤愚经》和《敦煌变文》中,也没有发现“於/于+单音节地点名词”结构,这点和Sun(2008)相符。在“时间”的左面子节点中,一共有8条数据带有单音节地点名词,其中5条不使用“於/于”。尽管这个现象说明在《春秋左传》中,单音节地点名词倾向单独出现,但是由于例子总数小,该结果在统计上的参考意义不是很大,我们对该结果持保守态度。

在“於/于”后音节的左面子节点中,首先由“之名词化”对包含多音节地点名词的数据进行递归分区。在其右面子节点“之名词化”中,一共有8条数据,其中5条使用了“於/于”。这个结果和Xu(2006)关于“於/于”在“之”名词化的名词前通常不省略的观点一致。但是同样的,由于例子总数较少,

对这个结果我们也应保持保守态度。

在“之名词化”的左面子节点中,“隐喻”是再次递归分区的分界因素。然而同上面一样,名词有隐喻含义的数据一共有8条,其中5条使用“於/于”,因此在统计上也不具大大参考价值。

另外两个因素“於/于”前音节和“文体修饰”没有出现在树形图中,这说明在统计上这两个因素和“於/于”的使用或省略不存在很大的关联。需要注意的是,尽管在收集数据过程中,存在一些於/于前单音节动词或名词带“於/于”的例子,但是整个数据中,也存在大量多音节带“於/于”的例子,所以“於/于”前音节只能算带“於/于”的充分但非必要条件。在和使用“於/于”的关联程度上,“於/于”前音节”也没有其他更为重要的因素高。同样地,当比较分析大量数据时,“文体修饰”在统计上的意义也变得很小。

5. 结论

从以上对递归分区结果的分析可以看出,在现代汉语之前,“於/于”的使用或省略首先与动词的及物性相关。高及物动词倾向于使用“於/于”来引出地点名词,而低及物动词则不是如此。汉语历史时期是第二个重要因素。从以上分析可知,“於/于”随着汉语的历史发展,尤其是在《水滸全传》以及之后的文献中,使用频率逐步减少,接近现代汉语(比如《儿女英雄传》)时,使用率接近零。但是这个现象并不说明现代汉语已经不需要地点介词,而是因为新的地点介词已经出现并取代了“於/于”。动词“死”是一个特殊现象,即使在早期现代汉语时期,也还是能大量带“於/于+地点名词”结构,关于这一点的原因还需进一步研究考察。下一个比较重要的因素是地点名词的音节数,单音节地点名词倾向于使用“於/于”。尽管这三个因素不能解释文献中的每一条数据,但是它们从统计角度出发,最大程度上解释了原始数据。除了这三个因素,之名词化和地点隐喻在统计上的重要性都相对比较较小,而“於/于”前音节”以及文体修饰的重要性则更小。

本文的统计分析结果和前人的观点可能存在不一致甚至相反的地方。然而它们是对一定数量的数据进行统计分析得出的,因此具有一定的统计学上的道理。前人的观点能各自解释一些数据,然而取样数量增大时,就出现了更多无法解释的例外。

综上所述,通过对历史语料的递归分区统计分析,本文从统计学的角度评估了所有和“於/于”使用相关的因素。本文的结果将作为进一步研究“於/于”使用,以及制约“於/于”使用的因素之间相互关系的基础。

附录 A

统计标记的例子

数据	於	时	及	“於/于” 后音节	之名词 化	“於/于” 前音节	地点 隐喻	文体 修饰
都淪死于水中	有	7	0	1	0	1	0	0
落下去淪死海中	无	5	0	1	0	1	0	0
攻殺周幽王于驪山之下	有	2	1	1	1	1	0	0
番陽人殺布茲鄉民田舍	无	2	1	1	0	0	0	0
楚、趙怒於魏之先已也	有	2	0	1	1	0	0	0
楚成王怒其不用其言	无	2	0	1	0	0	0	0
以公命取車於道	有	2	1	0	0	0	0	0
如何不取世間相也	无	4	1	1	0	0	1	0

附录 B

“於/于”数据递归分区的 R 编辑语言

```

> yu <- readtable("clipboard", header = T)
> library(rpart)
> fityu <- rpart(於 ~ 时间 + 及物 + “於/于”后音节 + 之名词化 + “於/于”
前音节 + 地点隐喻 + 文体修饰, data = yu)
> jpeg("D:/yu.jpg", height = 600, width = 600)
> plot(fityu)
> text(fityu, use.n=T)
> dev.off()

```

```

# fityu 结果
> fityu

```

n = 293

node), split, n, loss, yval, (yprob)

* denotes terminal node

- 1) root 293 106 有 (0.36117747 0.6382253)
- 2) 及物 < 0.5 172 80 有 (0.4651163 0.5348837)
- 4) 时间 < 4.5 127 52 无 (0.5905512 0.4094488)
- 8) “於/于”后音节 >= 0.5 79 21 无 (0.7341772 0.2658228)
- 16) 之名词化 < 0.5 71 16 无 (0.7746479 0.2253521)
- 32) 地点隐喻 < 0.5 63 11 无 (0.8253968 0.1746032) *
- 33) 地点隐喻 >= 0.5 8 3 有 (0.3750000 0.6250000) *
- 17) 之名词化 >= 0.5 8 3 有 (0.3750000 0.6250000) *
- 9) “於/于”后音节 < 0.5 48 17 有 (0.3541667 0.6458333)
- 18) 时间 < 1.5 8 3 无 (0.6250000 0.3750000) *
- 19) 时间 >= 1.5 40 12 有 (0.3000000 0.7000000) *
- 5) 时间 >= 4.5 45 5 有 (0.1111111 0.8888889) *
- 3) 及物 >= 0.5 121 26 有 (0.2148760 0.7851240) *

注释

1. 本文受到斯坦福大学孙朝奋 (Chaofen Sun) 教授和王靖宇 (John C. Y. Wang) 教授, 以及当时在斯坦福大学统计系做访问的 Michael Greenacre 教授的悉心指导, 谨致谢忱。本文一部分草稿在 2009 年第五届汉语语法化问题国际学术讨论会上 (上海师范大学) 作报告时, 也受到在场学者们的反馈意见, 特此表示感谢。
 2. 很多学者如闻春 (1985), Pullerblank (1995), 郭锡良 (1997) 等已经发现简体“于”在时间上出现在繁体“於”之前, 但是由于后者使用方法和意义比前者多, 因此在一些文献中直接取代前者。详细情况可以参见 Xu (2006)。本文的分析同时包括了作为地点介词使用的“于”和“於”。
- 本文的“古代汉语”是针对“现代汉语”而言, 根据汉语史研究一般做法, 将“五四”新文化运动作为古代汉语和现代汉语的分界点。“五四”

新文化运动以前, 汉民族共同使用的语言称“古代汉语”; “五四”新文化运动以后, 汉民族共同使用的语言称“现代汉语”。

3. 尽管《喻世明言》中的一些故事在《水浒传》之前就存在, 但是前者已被《喻世明言》的编者冯梦龙大量编辑过。冯梦龙是晚明清初人, 因此本文认为《水浒传》的语言早于《喻世明言》。

4. 搜索工作是在 www.sinica.stanford.edu 上完成的。该网站是台湾中央研究院建立的汉语历史文献语料库, 可供在线搜索。搜索时, 每个动词作为关键词输入, 之后人工寻找每条符合“动词+宾语/0+於/于/0+地点名词”结构的例子。“动词+宾语/0+於/于/0+地点名词”包括四种表层结构: “动词+地点名词”, “动词+宾语+地点名词”, “动词+於/于+地点名词”以及“动词+宾语+於/于+地点名词”。

5. 在古代汉语中, 很多低级动词比如“死”和“怒”都有使动用法, 从而带相应的宾语, 但本文考察过程中不包括此类用法。

6. 动词语义类型上的划分参考于 Levin (1993)。

7. P 值=0.05 是指在置信度 95% 的前提下接受假设的概率。若将原假设建立为数据符合正态分布, 如果 $P>0.05$, 则接受原假设, 即数据服从正态分布。反之则拒绝原假设, 即数据不服从正态分布。

8. 统计数据类型可以划分为三类: 分类数据, 顺序数据和连续性数据。分类数据是定性或品质数据, 比如我们可以把“及物性”定义为 0=低及物, 1=高及物, 2=双及物。且及物性与其代表的 0-2 的数值没有顺序上的制约关系, 也就是说, 我们也可以定义为 2=低及物, 1=高级物, 0=双及物。分类数据也采用数值代表, 但是和分类数据不一样的是, 其代表顺序的数值有计算上的意义。比如本文的文献是从七个不同汉语历史时段选取的, 因此可以把它们从最早期到最晚期排列并用 1-7 的数值标记。本文不涉及任何连续性数据。

9. 计算“及物”左面子节点使用“於/于”的方法: 将“及物”左面所有子节点末端“有於/于”的数字相加, 即 $(63-52)+5+5+(8-5)+28+40=90$ 。

10. 我们没有找到任何“取”在“动词+宾语+於/于+地点名词”结构中的例子, 而“杀”在《喻世明言》中出现过一次“动词+宾语+於/于+地点名词”的例子: “諸侯皆不來救, 犬戎遂殺幽王於驪山之下”。该例子受古文影响可能性比较大, 因为在《史记》中可以找到类似的例子: “申侯怒而與犬戎共攻殺周幽王于驪山之下”。

引用参考文献

- 董秀芳. 2006. 古汉语中动名之间“于/於”的功能再认识. 《古汉语研究》71:2-8.
- 方平权. 2000. 关于介词“于”由先秦到汉发展变化的两种结论. 《古汉语研究》47:70-73.
- 郭锡良. 1997. 介词“于”的起源和发展. 《中国语文》2:131-138.
- 何乐士. 1985. 《左传》, 《史记》介宾短语位置的比较. 《语言研究》1:57-65.
- 时兵. 2003. 也论介词“于”的起源和发展. 《中国语文》4:343-347.
- 闻宥. 1985. “于”“於”新论. 《中国语言学报》2:44-48.
- 吴国忠. 1989. 古汉语介词“于”使用的随意性. 《求是学刊》6:75-79.
- 许敏云. 2008. 试析《史记》中的“于”, “以”结构. 《南昌高专学报》74:89-90.
- BAAYEN, R. Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- HOPPER, J. Paul, and Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2):251-299.
- LAMBERT, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representation of Discourse Referents*. Cambridge: Cambridge University Press.
- LEVIN, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
- PULLEYBLANK, Edwin. 1995. *Outline of Classical Chinese Grammar*. Vancouver: UBC Press.
- RAPPAPORT HOVAVY, Malka, and Beth Levin. 2008. The English dative alternation: The case for verb sensitivity. *Linguistics* 44:129-167.
- SUN, Chaofen. 2006. *Chinese: A Linguistic Introduction*. Cambridge: Cambridge University Press.
- _____. 2008. Two conditions and grammaticalization of the Chinese locative. In *Space in Languages of China Cross-linguistic, Synchronic and Diachronic Perspectives*, ed. by Dan Xu. Dordrecht: Springer.
- XU, Dan. 2006. *Typological Change in Chinese Syntax*. Oxford: Oxford University Press.
- ZHANG, Heping, and Burton H. Singer. 1999. *Recursive Partitioning and Applications*. New York: Springer.

FACTORS RESTRICTING THE USE OF
THE LOCATIVE PREPOSITION *yu* IN CLASSICAL CHINESE:
A MODELING OF RECURSIVE PARTITIONING

Jingxia Lin

Nanyang Technological University, Singapore

ABSTRACT

Different accounts (Fang 2000; Dong 2006; Xu 2006; Xu 2008; et al.) of the optional use of locative preposition *yu* (於/于) in classical Chinese have been proposed. Yet there is still no agreement. This project proposed a new account via a statistical modeling method known as “recursive partitioning” to identify all the relevant factors correlated to the optional use of *yu* proposed previously. It is found that *yu* tends to follow verbs with high transitivity, which is opposite to Dong (2006) and Xu (2006). The historical development of Chinese is found to be the second most significant constraint: *yu* is less frequently used especially after the time of *Shuihu Quanzhuan*. In addition, *yu* tends to occur with a following monosyllabic NP. Other factors such as focus, *zhi* (之)-nominalization and prosody are found statistically insignificant. The results of this paper will serve as a quantitative base for further research on the grammatical function of *yu*.

SUBJECT KEYWORDS

Locative preposition *yu* Modeling of recursive partitioning