九州大学学術情報リポジトリ
Kyushu University Institutional Repository

计量文体分析和戏剧的作者认定

太田, 一昭
九州大学大学院言語文化研究院 : 学術研究者

廖, 柳棉
華東師範大学外語学院

https://doi.org/10.15017/2560389

出版情報：言語文化論究. 44, pp.67-74, 2020-03-13. 九州大学大学院言語文化研究院
バージョン：
権利関係：

计量文体分析和戏剧的作者认定

太田　一昭（著）・廖 柳 棉*（译）

1．Shakespeare 写了多少部戏剧呢

先来一个简单的问题，William Shakespeare（1564-1616）在一生中写过多少部戏剧呢？我虽然说“简单”，但其实要回答这个问题却并不简单。在过去，要回答是很容易的，只要回答“37部”就好了。1623年 Shakespeare 的戏剧全集被首次出版，这部全集（以“两折本”这种大型开数出版）收录了36部戏剧，这36部之外再加上 *Pericles*（在1609年以“四折本”的较小型开数首次出版）这部被认为出自 Shakespeare 之手的作品，共计37部戏剧。

（1）Shakespeare 戏剧全集（1623年）收录36部＋ *Pericles* 共计37部

这37部长篇巨著，都是被看作是 Shakespeare 戏剧的“真作”。这些“真作”的范围，从20世纪后半叶开始，就有逐渐扩大的趋势。所以如今37部未必是正确的答案，以下解答可供参考。

（2）38部（36部＋ *Pericles* ＋ *The Two Noble Kinsmen*）
（3）39部（36部＋ *Pericles* ＋ *The Two Noble Kinsmen* ＋ *Sir Thomas More*）
（4）40部（36部＋ *Pericles* ＋ *The Two Noble Kinsmen* ＋ *Sir Thomas More* ＋ *Edward III*）
（5）41部（36部 ＋ *Pericles* ＋ *The Two Noble Kinsmen* ＋ *Sir Thomas More* ＋ *Edward III* ＋ *Double Falsehood*［*Cardenio*］）
（6）42部（36部 ＋ *Pericles* ＋ *The Two Noble Kinsmen* ＋ *Sir Thomas More* ＋ *Edward III* ＋ *Double Falsehood*［*Cardenio*］ ＋ *Arden of Faversham*）

上述标题中出现的戏剧，全部都是 Shakespeare 和其他剧作家共同完成的“合著”。可能有不宜把“合著”列为 Shakespeare 作品的看法，但如果只把单独执笔的作品算作 Shakespeare 戏剧的话，Shakespeare 戏剧就只有27部 。（根据研究者的不同数字会有些许差异。也有研究者认为 Shakespeare 单独执笔的作品有28部）。也就是说，37部“真作”中包含着“合著”。因此，把“合著”看作是 Shakespeare 戏剧是没有问题的。问题是，那些戏剧究竟是不是 Shakespeare 本人写的。

* 华东师范大学外语学院研究生

２．Shakespeare 的“真作”的扩大和 *The New Oxford Shakespeare*（2016）

Shakespeare 的作品集（全集和分册）几乎被刊发了无数次，在2016年刊发的 *The New Oxford Shakespeare: The Complete Works*（以下略为 *NOS*），是主流 Shakespeare 全集丛书中，最热衷于“真作”扩大的一版。这部 Shakespeare 全集如今引起了强烈的争议，议论的焦点，是它与其说全新不如说是激进的编辑方式。*NOS* 中，*The Two Noble Kinsmen* 自是不必说，还收编了 *Thomas More*、*Edward III*、*The History of Cardenio*［*Double Falsehood*］、*Arden of Faversham* 和 *Spanish Tragedy*（1602年加上了台词）收录 *The Two Noble Kinsmen* 是一直以来的主要的Shakespeare全集的惯例。*Sir Thomas More* 和 *Edward III* 已经被收录在 *The Riverside Shakespeare* 全集的第二版（1997）和 *The Oxford Shakespeare: The Complete Works* 第二版（2005）中。以及 *The History of Cardenio*，也就是 *Double Falsehood* 也已作为 The Arden Shakespeare, 3 rd ser 丛书的一部分（2010）出版了。然而 *Arden of Faversham* 和 *Spanish Tragedy*（附加台词）被收录进主要的 Shakespeare 全集中还是头一次。在前面已经说过，20世纪后半叶之后，Shakespeare“真作”的认定范围呈现扩大的趋势，但现在，*The Two Noble Kinsmen* 自是不必说，不反对把 *Edward III* 作为 Shakespeare 的“合著”入 Shakespeare 的“真作”之中的学者也多了起来。然而连 *Arden of Faversham* 也算作“真作”，甚至认为 *Spanish Tragedy* 的附加部分也是 Shakespeare 所写，*NOS* 如此盖章认证着实令人吃惊。*NOS* 做的不止是扩大 Shakespeare 的“真作”范围，还明确地区分了一直以来多被怀疑是合著的戏剧的作者。比如 *NOS* 认定，*Henry VI, Part 1* 是 Shakespeare 改编自 Christopher Marlowe 和 Thomas Nashe 以及另一位作家的合著；*Henry VI, Part 2* 和 *Henry VI, Part 3* 是 Shakespeare 改写自己和 Marlowe 及另一名作者的共同作品而成。对 Shakespeare 的合著的共同作者的认定，*NOS* 有着长久的研究积累。*NOS* 的编辑和学者们自然是根据之前的研究成果进行的作者断定，但 *NOS* 的共同作者断定却比前人具体且大胆。而支持这样大胆的断定的，是利用计算机进行的统计学作者认定。

３．*NOS* 的基于统计学的作者断定方法

NOS 的编辑和学者们的作者断定方法，在 *The New Oxford Shakespeare: Authorship Companion*（2017）里有详细的介绍。*NOS* 的编辑学者们使用的统计学分析方法涉及多个方面，虽然多通过计算机程序来收集和分析语料，仍有部分语料要用从前就一直进行的“手工操作”来分析。但在人工分析语料时，也会利用计算机来收集和整理数据。利用计算机来进行统计学上的分析，这种方法对于“外行”来说不难理解，但要评价分析结果是否得当却不是那么简单的事情。Delta、Nearest Shrunken Centroid、Random Forests、Zeta、Principal Component Analysis、Shannon Entropy —— 这些是 *NOS* 的编辑学者们用来进行作者断定的部分分析方法。除此之外，还使用了比较稀少词汇的出现频率、音律的作家间比较等较为传统的作者断定方法，但大多数还是对普通文学研究者来说理解困难的、多变量分析等计量文体学分析方法。*NOS* 的编辑学者们似乎在宣扬已经可以根据统计学分析来完成作者的认定。

NOS 编辑们（利用“外行”难以反驳的分析方法来使自己的作者断定具有权威性）的这种态度，也引来了一些批评。Shakespeare 的自由研究者 Pervez Rizvi 是批判的急先锋，Rizvi 的个人简历因为不公开所以不甚清楚，但他是一名精通计算机编程和统计学的数学家。Shakespeare 的相关研究，以、“Evidence of Revision in Othello,” *Notes and Queries* 45.3（1998）：338-343为始，文献学相关的论文少说也有10篇已经被公开刊登，其中多数与作者认定有关。Rizvi 对 *NOS* 的基于统计学的作者统一性认定方法进行了强烈的批判，Rizvi 认为，*NOS* 的编辑们的计量文体分析方法纯属杜撰，展开的作者认

定的讨论也没有说服力。他们抓住文学研究者通常不熟悉由计算机进行的文体分析这一点，借着的歪门邪道的统计处理方法给自己的原书编纂制造噱头，但实际上是连简单的统计分析方法都不懂的无能之辈。

我虽然不像 Rizvi 一样拥有统计学素养和数学知识，但我觉得 Rizvi 的批评是对的。*NOS* 的编辑们认为已经用计算机计量文体分析解决了 Shakespeare 的“合著”问题，计量文体分析仿佛是解决作者认定这一疑难问题的魔法棒。然而我认为至少在现阶段，用计算机计量文体分析来确定 Shakespeare 的合著作者不过是个传说。在今天的报告上，我将叙述这样想的原因。

批判 *NOS* 的 Rizvi 在2017到2018年间，在网站上公开了初期近代英国戏剧惯用语数据库。该数据库把527部初期近代的英国戏剧（将出版于1602年的 *The Spanish Tragedy* 的附加台词作为独立的一部作品计入）中出现的惯用语相互进行了一致性匹配。Rizvi 建立的这个数据库，就笔者所知，是研究初期近代英国戏剧作者认定的历来最有力的资料。Rizvi 的惯用语一致数据中，附有收集总结了各戏剧的数据的文件，和（Rizvi 自己编写的）分析这些数据的作者认定程序。我在验证了 Rizvi 的数据库和作者认定程序可以辨别作者到何种程度之后，认为它可以成为断定作者的一个强有力的工具（后文将就此详述）。还明白了，这也能成为证明利用计量文体分析来断定作者一事之困难的间接证据。于是今日我想通过 Rizvi 的这份资料，来指出几个 *NOS* 基于统计学的作者断定方法的疑点。在正式开始之前，先以 *Arden of Faversham*（推测初次演出为1591年）的作者认定方法为例，给大家具体展示力推 *NOS* 的编纂方法的 *NOS* 团队是如何进行作者断定的。

4．通过计量文体分析进行的 *Arden of Faversham* 的作者认定

Arden of Faversham 是一出作者不详的“家庭悲剧”，有位叫 Thomas Arden 的乡村地主，他的妻子和妻子的情人一起合谋杀掉了他。Thomas Kyd、Marlowe、Shakespeare 等人被视为此剧可能的作者，但一直以来只有少数学者主张把 Shakespeare 作为该剧作者的有力人选。

Authorship Companion 里有两篇关于 *Arden of Faversham* 的作者认定的考证，其中一篇是由两位学者（Elliott and Greatley-Hirisch）共同完成的。他们先调查了 9 位剧作家的（推测在1587年至1594间初次演出的剧目）共计34部戏剧中全部单词的出现频率，或者功能词的出现频率，再或500个高频词的出现频率等，再形成了各个作家的“档案”（表示各作家特点的数字化语料）另一方面，他们把 *Arden of Faversham* 分成35个“rolling segments”，再从中抽取相同的语言数据。第一片段的语料数量是2000词。（“rolling segments”指的是包含和其他片段重复单词的片段集合。第一片段有 1 -2000个词，第二片段有501-2500个词，第三片段有1001-3000个词，由此可分出35个片段。）接下来把各片段中的数据和 9 位剧作家的“档案”进行比较，从而推测出该片段的作者。此时分析人员们用的统计学断定方法叫做 Delta、Nearest Shrunken Centroid、Random Forests。*NOS* 的编辑学者们再抽出500个特定作家的 marker words（作者标记词），检查 *Arden of Faversham* 的35个片段中，根据上述断定方法确定的可能的 5 位作者的戏剧样本和片段集合中 marker words 出现的频率，再根据这一频率来推测 *Arden of Faversham* 的35个片段的作者。进而把 *Arden of Faversham* 的可能作者最后锁定到 3 位（Kyd、Marlowe、Shakespeare），再用上述的语料和分析方法（包括使用500高频词和功能词出现频率数据的主要成分分析）进行再分析，从而做出最后的判断。根据这个判断，*Arden of Faversham* 不是由 Shakespeare 独立完成（以 Shakespeare 为主要的作者），而是 Shakespeare 和其他作家合写的作品。

另一篇考证的作者是 M. P. Jackson，Jackson 的分析立足于计算机计量文体分析的研究成果，他自己现在通过计算机收集、计算和整理数据，但既没有使用复杂的统计学方法，也不是用计算机程序来

进行统计分析。他用的方法，更接近沿用至今的所谓人工统计分析法。Jackson 先根据先行研究，筛选出了 5 个 Shakespeare 标记词（Shakespeare-plus-words）和 4 个非 Shakespeare 标记词（Shakespeare-minus-words），再检查1580年到1600年间被推测为 Shakespeare 所写的（现存的）全部137部戏剧中，两种标记词的出现频率，最后计算出各部戏剧中 Shakespeare 标记词的相对出现率。出现率用 Shakespeare 标记词出现的数量，除以 Shakespeare 标记词的出现数和非 Shakespeare 标记词的出现数之和，再乘上100的数值表示（百分比）。由此可知，Shakespeare 剧总体而言，会显示比非 Shakespeare 的作家所作的戏剧更高的数值。加上这个数据分析结果，同时根据（出现频率相对较小的）惯用语的 *Arden of Faversham* 和其他戏剧对比的一致性数据，可以下结论说，*Arden of Faversham* 是 Shakespeare 的“合著”，尤其从第 4 场到第 9 场的部分，是 Shakespeare 所写的可能性很大。

5．*NOS* 的基于统计学的作者认定方法的疑点

NOS 团队的 *Arden of Faversham* 的作者断定方法大致如上。如果是对计量文体学和统计学上的作者认定研究熟悉的研究者，那么经过如上说明应该可以充分理解，但对普通的文学研究者来说要理解可能还是有难度。其实就连我自己，也有很多不能理解的地方。比如 Elliott=Greatley-Hirisch 是经过怎样的步骤来运用 Random Forests 来进行作者断定的呢？我虽然理解了分析方法的原理，但因为没有提供具体数据，仍然无法完全理解。但是，能在我今日的报告中，把握住 *NOS* 的编辑学者们究竟用了何种方法来尝试断定作者的大体印象就已经足够了。*NOS* 团队断定 Shakespeare“合著”作者的方法虽然会根据作品的不同或多或少有差异，毫无疑问的是他们打着计量文体分析下的作者的旗号，想要给给自己大胆的编纂方式树立权威。我抱有怀疑的，是 *NOS* 的作者认定方法。在这里不可能指出所有的疑问点，谨就几个重要的地方论述如下。

如前所述，*NOS* 的学者们运用了各种各样的分析方法，似乎多种、大量的分析方法产生的分析结果都支持了他们作者认定的正确性。然而，就算进行了多重分析，也未必能做出更好的作者认定。更重要的是选择恰当的分析材料。不论多出色的分析方法，只要作为分析对象的语料有偏差，就无法得出值得信赖的分析结果，这就是 *NOS* 团队的分析所存在的最大问题。

Elliott=Greatley-Hirisch 从作为分析基础资料的 9 位剧作家共34部戏剧中抽取了语言数据，他们判断 Shakespeare 是 *Arden of Faversham* 的最有可能的作者人选，但作为判断根据的数据却是从 Shakespeare 早期的 4 部作品中抽取出来的。就是 *The Two Gentlemen of Verona*（1590）、*The Taming of the Shrew*（1591）、*Richard the Third*（1592）、*The Comedy of Errors*（1594）这四部。（括号内的数字是推测的初次上演年份）乍一眼看，分析者们选择这 4 部作品似乎是合理的，因为这 4 部剧都是和 *Arden of Faversham* 几乎在同一时期完成的作品。同时期内 Shakespeare 写的其他作品中，*Henry VI* 三部曲广为人知。然而 *Henry VI* 三部曲很可能是 Shakespeare 和其他剧作家共同写作的作品，所以分析者们没有从 *Henry VI* 中抽出语料形成 Shakespeare 的“档案”也是可以理解的。但他们选用的 4 部作品是否能代表 *Arden of Faversham* 在创作首演时的 Shakespeare 的语言特点，这一点实际上是无从得知的。我还关注的一点是，用来比较对照的戏剧的种类差异。*Richard III* 是历史剧，而其他的作品是喜剧，一方面又把 *Arden of Faversham* 分为“家庭悲剧”。究竟能不能通过这样的分类来确认 *Arden of Faversham* 是否和 Shakespeare 有关系，不免心存疑惑。

剩下的 8 位剧作家的戏剧样本，他们现存的单独写作的戏剧几乎都被选为比较对照的文本，但是根据作家的不同其数量也并不平均。John Lyly 有 8 部，Marlowe 有 5 部（不包括 *Dr. Faustus*），Kyd 有 3 部（并非 Kyd 的原创戏剧，包括 Kyd 翻译成英语名为 *Cornelia* 的 Robert Garnier 的法语作品），Robert

Greene 有 4 部，George Peele 有 5 部，Robert Wilson 有 3 部，Thomas Lodge、Thomas Nashe 各有 1 部。

如上，根据 Shakespeare 和其他剧作家合计 9 位剧作家的34部戏剧形成的语料库数据，他将 *Arden of Faversham* 的作者（或者说主要作者）断定为 Shakespeare。但至少从语料的组成来看，我认为不能全面信任这样的分析结果。把仅从 1 部戏剧中得出的作家的数据作为分析的基础资料，不妥当之处自是不必说，就算从几部作品中抽取出来的数据，也未必就能正确表示该作家的特点。之所以这么说，是因为若是戏剧样本数量少，根据样本的选择，分析的结果很可能会产生大的变动。

Elliott=Greatley-Hirisch 的分析中，并没有公开分析时使用的原始数据，因此第三方无法检验分析结果。分析结果的一部分由图表，或者说分布图来表示。虽然分析者们自信满满地下了结论，但我看来，他们的分析结果并不像他们所声称的那样能够推导出明确的结论。关于这一点，由于时间关系无法详述，但可以说的是，出现了根据分析的戏剧样本的选择差异，作者判断也产生了些许变化，这种结果值得深思。

Jackson 的分析中，用了比 Elliott=Greatley-Hirisch 更多的戏剧数据来断定，足足有137部，其中包括23部 Shakespeare 的戏剧（单独完成的作品和合著）和29部作者不详的作品。Jackson 重视 Shakespeare 标记词的出现率，最重要的分界点（Borderline A）（超过此点则为误判少）是68.5％。被认为很可能是 Shakespeare 所写的第 4 场到第 9 场的出现率平均是86.7％，其他部分平均是57.9％。因此断定，第 4 场～第 9 场为 Shakespeare 作品的可能性很高。

Jackson 的分析存在几点问题。虽然在计量文体分析中戏剧样本的数量多多益善，但我认为 Jackson 是为了让自己的结论自圆其说，而只使用了对自己有利的数据。Elliott=Greatley-Hirisch 也使用了 Jackson 用过的 Shakespeare 的戏剧样本，包括（Shakespeare 初期的） 4 部戏剧。这 4 部戏剧的 Shakespeare 标记词的出现率如下：*Richard III* 59.5％，*The Two Gentlemen of Verona* 64.9％，*The Comedy of Errors* 59.5％，*The Taming of the Shrew* 59.0％。若是存在这种分布的不均匀，根据平均值做的判断也是不合理的。特意去求这 4 部戏剧的平均值的话，结果则是66.8％。也就是根据戏剧样本选择的不同，第 4 场～第 9 场的数值就初期的（也就是 *Arden of Faversham* 创作首演时）Shakespeare 来说，未免过高了。

Jackson 引用了 Hugh Craig and Arthur F. Kinney, eds. *Shakespeare, Computers, and the Mystery of Authorship*（2009）来为自己的理论背书，Craig=Kinney 和 Jackson 同样宣扬 *Arden of Faversham* 是 Shakespeare 的合著这一观点，但 Jackson 的部分分析数据和 Craig=Kinney 有矛盾之处。Jackson（p. 193）表示，第 4 场～第 9 场清楚地是 Shakespeare 的作品。尤其第 8 场是 Shakespeare 所写的可能性非常大（p. 190），但 Craig=Kinney（p. 93）的分析结果（基于 Shakespeare 标记词和非 Shakespeare 标记词出现频率的 Zeta 检测结果）显示它是除 Shakespeare 之外的作家所写的可能性很大。如上所示，Jackson 并未提及 Craig=Kinney 的分析和自己的分析结果相矛盾这一点。一方面，Jackson 引用的 Craig=Kinney 的文章也存在着问题。Craig=Kinney 得出的最终结论和 Jackson 一样，都表示第 4 场～第 9 场为 Shakespeare 所作的可能性非常大，但我认为 Craig=Kinney 是在为了让自己的分析结果符合结论而强行解释。这是因为虽然 Craig=Kinney 的数据没有和他们的结论产生矛盾，但论证过程却没有说服力。（顺便一说，根据 Jackson 的分析方法，无法得出可以“清楚地”断定第 9 场 是 Shakespeare 作品的数据。根据我的调查，Shakespeare 标记词的出现率是60％。也就是该百分比在 Borderline A 之下）

Elliott=Greatley-Hirisch 的分析结果也和 Jackson 所主张的不同，Elliott=Greatley-Hirisch 认为还不能“清楚地”显示第 8 场的作者就是 Shakespeare。Elliott=Greatley-Hirisch 的分析结果（p.180）却相反地显示了，在 Jackson 的分析中被认为不是 Shakespeare 作品的场次（除第 4 场合第 9 场之外的场次）大部分都是 Shakespeare 的作品。Elliott=Greatley-Hirisch 和 Jackson 作为 *NOS* 团队的学者，应

该都意识到自己的分析结果相互存在不一致的地方，然而他们却都对这样的龃龉缄口不提。

以上，就 *NOS* 的作者认定的疑点，以 *Arden of Faversham* 的分析为例进行了说明，其他作品的分析也存在相同的问题，*NOS* 的编辑学者们却把计量文体分析当做对作者认定问题的终极解决方案展开了论证。然而想靠计量文体分析来解决作者认定问题，是痴人说梦。我们至少在现在这个阶段有必要认识到，运用计量文体分析来进行明确的 Shakespeare 时代佚名戏剧的作者认定是做不到的。接下来，我想就此（明确地断定作者是很困难的事），用之前介绍的 Rizvi 的数据库和作者验证程序来举例证明。

６．Pervez Rizvi 的戏剧间惯用语一致数据库和作者认定

Rizvi 的数据库里的数据，不是527部作品的词频数据，而是作品间的语句一致性数据。也就是调查出现在某部戏剧和其他526部戏剧中全部一致的惯用语（短语），再记录、整理和统计该数量后得出的值。虽然过去不把作品间的语句一致性作为表示作者同一性的标志，但 Rizvi 用（他自己编写的）计算机程序来进行了该项调查，之后在线上公开了详细的调查结果。尽管使用 Rizvi 数据库的英国文学研究者（至少据我所知）在日本几乎没有，但这个数据库是对作者认定而言非常宝贵的工具。我用 Rizvi 数据库试着验证了它的辨别作者的程度，验证的具体过程和结果另行说明，此文着重探讨经过此番验证得出的看法。如下所示：

（1）Rizvi 的数据库是有着高精度的戏剧作者认定工具。
（2）在断定词数少的片段（短小的场景）的作者时，精度可能不高。这是因为通常情况下，词数少的片段的粗频率数据（戏剧间惯用语一致数）会变小，易导致基于粗频率小数据的分析结果产生误差。
（3）大部分作家的作品中，都包含和语料库的其他语料相比“异常”的语言特征的作品。只要有“异常”的作品被包含在该作家的语料库里的可能性，用计量文体分析来完美地进行作者认定就很困难。
（4）就算是同一个作家，也不应该采集不同类别的作品数据。要做戏剧的作者认定，基础资料就应该只从戏剧当中抽取。比如有一位作家普通的戏剧和假面剧都写，那么假面剧就应该从断定作者的语料库中删除。
（5）在断定佚名作品的作者时，可能作者的数据最好从和判断对象的创作年份相近时期写就的作品当中抽取数据。这是因为，一个多年从事创作活动的作家，他的写作风格和语言特征有可能发生变化。
（6）在根据少量作品的数据判断作者时，可信度会下降，因为作品选择的不同易导致分析结果发生变化。
（7）*The Taming of the Shrew* 有除了 Shakespeare 之外的人所写的可能性，因此它不应该作为判断是否为 Shakespeare 所写作品的基础资料。

以上的验证结果在表示、证明用计量文体分析来判断作者有一定有效性的同时，也显示了完美的统计学作者断定是不可能的。同时还揭示了 *NOS* 的 Elliott=Greatley-Hirisch 主张自己成功判断了 *Arden of Faversham* 的作者分析有多脆弱。从仅有的几部戏剧样本中得到作家数据，再基于这种数据进行的分析，不论用了多厉害的统计学断定方法，都无法由此得到可信的结果。当然我自己的验证也是基于 Rizvi 的数据库和作者认定程序得出的，因此可能会有意见认为，从这一结果推测其他的计量文体分析存在的问题未必妥当。然而，从词数少的文本中抽取出来的数据，和基于粗频率数据较小的资料所进行的分析，它们都容易受影响而变化且可信度不高，这一点对统计学者来说是常识。以及一直以来，都在强调用来比较对照的资料，应该从和作者断定对象的作品相近年份的文本中抽取的重要性。这就是说，不论是词频数据分析还是惯用语一致性数据分析，统计学上的作者认定方法都面临着共同的问

题。若是如此，不论采用了何种方法，存在上述问题的计量文体分析的断定结果都不能被信赖。

7．结语——统计学上的作者断定的界限

如上，我指出了 *NOS* 的统计学上的作者认定方法的问题之处，但要确切地用统计学分析来认定 Shakespeare 时代佚名作品的作者还是有难度。一般来说，戏剧文本会暗含着不稳定因素，它存在于从脚本的完成到版本的印刷的过程中，除了作者之外的工作人员（剧团、排字工人、印刷人员等）的参与可导致和原始文稿的差异，作者的认定也因此变得困难。加之那个时代的戏剧多为合著，很难保证计量文体分析所必要的、足量的独立写作而成的文本。Shakespeare 因为有27部（合计62万字）的独立写作的作品，所以可以抽取出较为稳定的语言数据，但 Shakespeare 同时代的作家中，没有人可以确保有这么多语料形成语料库。比如 Marlowe 大约有10万词，Lyly 大约有11万词。Kyd 的语料库大约有 4 万词，但其中包括了 Kyd 的（非原创戏剧）英译戏剧和（没有明确证据证明为 Kyd 独立写作的）*Soliman and Perseda*。把仅有的 1 部作品作为该作者的语料库的例子并不鲜见，也存在很多作者不详的戏剧。虽然 Shakespeare 可以确保27部独立写就的文本基本是个例外，但不清楚那些文本是否全部是 Shakespeare 的真迹，因为戏剧的剧本有可能包含被剧团的相关人员改动过的台词。以及 Shakespeare 初期（和晚年）有许多合著，要确保足量的分析基础资料也是困难的。这就是难以用计量问题分析来判断 Shakespeare 的初期作品，和同时期写就的 *Edward III*、*Arden of Faversham* 和 Shakespeare 之间的关系的另一个原因。况且还存在与作者的通用语法截然不同的“异常”作品，若是这种情况，不论如何使用高辨别度的分析方法，计量文体分析对作者认定能做的无疑有限。不得不说，*NOS* 团队这种无视界限，或者说仿佛界限不存在一样的作者断定方法是有问题的。

引 用 文 献

Craig, Hugh, and Arthur F. Kinney, eds. *Shakespeare, Computers, and the Mystery of Authorship.* Cambridge: Cambridge UP, 2009.

Elliott, Jack, and Brett Greatley-Hirsch. “*Arden of Faversham*, Shakespearean Authorship, and ‘The Print of Many’.” Taylor and Egan, *Authorship Companion* 139-81.

Jackson, M. P. *Determining the Shakespeare Canon: “Arden of Faversham” and “A Lover’s Complaint.”* Oxford: Oxford UP, 2014.

———. “Shakespeare and the Quarrel Scene in *Arden of Faversham*.” *Shakespeare Quarterly* 57.3 (2006): 249-93.

———. “A Supplementary Lexical Test for *Arden of Faversham*.” Taylor and Egan, *Authorship Companion* 182-93.

Kinney, Arthur F. “Authoring *Arden of Faversham*.” Craig and Kinney 78-99.

Rizvi, Pervez. “Attribution Studies Article.” Shakesper, 8 Sep. 2018. <https://www.shaksper.net/archive/2018/827-september/32577-pervez-rizvi-attribution-studies-article-4>.

———. “Collocations and N-grams.” Published online: 30 May 2018. Web 30 July 2018. <http://www.shakespearestext.com/can/index.htm>.

———. “Small Samples and the Perils of Authorship Attribution for Acts and Scenes.” *A Quarterly Journal of Short Articles, Notes and Reviews*. Published online: 13 Nov 2018. Advance Access Article. Web. 1

December 2019. <https://www.tandfonline.com/doi/full/10.1080/0895769X.2018.1537841>.

Taylor, Gary, and Gabriel Egan, eds. *The New Oxford Shakespeare: Authorship Companion*. Oxford UP, 2017.

Taylor, Gary, John Jowett, Terri Bourus, and Gabriel Egan, eds. *The New Oxford Shakespeare: The Complete Works*. Modern Critical Edition. Oxford: Oxford UP, 2016.

Taylor, Gary, and Rory Loughnane. "The Canon and Chronology of Shakespeare's Works." Taylor and Egan, *Authorship Companion* 417–602.

追记：

本文原载于《首届"戲單、劇場與二十世紀上半葉的東亞演劇"學術研討會論文集》（中里見敬編，九州大學言語文化研究院，2019年 8 月）。为此次发表进行了一些修订。

本研究得到 JSPS 科研费 JP16H03405 的经费赞助。