arXiv:2302.03645v2 [cs.CL] 8 Feb 2023

文本演化中的利用与探索。量化写作过程中的规划与翻译流。

D. Ruggiero Lo Sardo[1,2,3,*], Pietro Gravino[4,1,2], Christine Cuskley[5], Vittorio Loreto[4,1,2,6],

1 索尼计算机科学实验室罗马分部，CREF-SONY联合倡议，恩里科·费米研究中心，Via Panisperna 89/A，00184，罗马，意大利**2** 恩里科·费米研究中心，Via Panisperna 89/A，00184，罗马，意大利**3** 复杂性科学中心，维也纳，A-1080，维也纳，奥地利**4** 索尼计算机科学实验室巴黎分部，6, Rue Amyot，75005，巴黎，法国**5** 语言演变、习得与发展研究组，纽卡斯尔大学，泰恩河畔纽卡斯尔，NE1 7RU，英国**6** 罗马智慧大学物理系，Piazzale Aldo Moro 2，00185，罗马，意大利

* donaldruggiero.losardo@sony.com

摘要

写作是一个复杂的过程，是现代人类活动中许多事务的核心。尽管写作看起来是线性过程，实则隐藏着许多高度非线性的过程。以往研究主要关注写作的三个阶段：规划、翻译与转写，以及修订。研究虽已表明这些阶段具有非线性特征，但在测量时往往仍将其视为线性过程。本文提出了一些测量方法，用于检测并量化写作过程中规划（探索）和翻译（利用）的子循环。我们将这些方法应用于一个全新的数据集，该数据集记录了文本创作的所有阶段，从早期尝试到最终版本的最后润色。该数据集来自一系列写作工作坊；借助创新的版本管理软件，我们得以记录文本构建过程中的每一个步骤。60多名科学领域的初级研究人员撰写了一篇面向普通读者的科学短文。我们将每篇文章记录为一个写作云，并将其定义为一种复杂的拓扑结构，用以捕捉文章本身的历史。通过这套独特的写作云数据集，我们呈现了写作过程的一种表征，量化了其复杂度，以及作者在草稿创作过程中随时间付出的努力。有趣的是，这种表征凸显了“翻译流”阶段——作者在此改进已有观点；也凸显了探索阶段——作者回到规划阶段时会出现创造性偏离。随着写作过程推进、作者逐渐接近最终版本，翻译与探索之间的这些转折点会变得越来越少。我们的结果以及提出的新测量方法，有望推动关于写作非线性本质的讨论，并助力开发支持更具创造性和影响力的写作过程的工具。

作者总结

写作是一个复杂的过程，包括组织思想（规划与探索）、将其转化为文本（翻译），最后修订并润色文本。尽管

读者是线性地阅读文本的，但生成文本背后的认知过程并非线性。写作中的非线性方面远未得到充分理解，中间版本的起伏变化也同样如此。在这里，我们通过记录文本从早期尝试到最后润色过程中不同版本的快照，揭示这些过程。我们提出了一些衡量指标，用以展现这一过程的复杂度，并检测写作过程中的探索性尝试、作者效率特别高时的心流状态，以及二者之间的转换。我们的结果为写作科学及其背后的创造过程提供了新的见解。

引言

在以读写为主的社会中，有效阅读和写作的能力是知识经济中个人必不可少的一项能力。读写能力是衡量幸福感的诸多指标（例如自评健康状况、信任、收入和融入程度）的制约因素 [1]。任何着手以书面形式进行交流的人，都会面临诸多挑战，例如构建逻辑论证、编织易于理解且引人入胜的叙事，以及做出恰当的词汇和语法选择。写作者如何将自己的想法组织成他人能够阅读、理解，甚至可能乐于阅读的叙事？心理学、语言学和教育学领域积累了大量理论和实验研究，并据此建立了广受认可的写作过程模型。然而，针对自然情境数据的研究仍然不足；这类数据不仅记录最终成品，也记录写作的过程 。本文借助记录写作全过程的版本管理软件，提出一个框架，依据这些模型分析写作过程中的自然情境写作。首先，我们概述认知科学对写作过程的基本理解。我们指出，尽管写作过程的各个阶段被认为并非线性推进、而是会相互反馈，但在方法论上，它们往往被视为彼此离散的实体。为此，我们提出一种分析创作完整写作成品全过程的方法，重点关注文本本身所产生、能够表明阶段转换的信号。

理解写作

认知科学家发现，写作过程包含三个阶段，各阶段涉及不同的子任务：规划、翻译和转录（句子生成），以及修订 [2]。规划阶段包括确定一篇作品的总体目标，并组织最初的想法。翻译和转录阶段是将这些想法转化为具体语言形式的过程，其中包括将想法调入工作记忆（翻译）、激活正字法知识，以及调动参与手写或打字的运动区（转录）。最后，修订阶段包括评估并反复评估文本，然后进行修改。重要的是，这一过程从根本上说并非线性的，因为修订阶段包含反复进行规划和翻译的循环：这些阶段从根本上彼此交织 [2]。修订行为通常涉及调整最初的计划，并通过生成新句子或改写句子来落实这些调整。换句话说，这些阶段“并非严格按序进行；相反， [它们]会递归地相互作用.....彼此打断，并相互嵌套” [36]。

教育学和心理学领域的研究者采用了多种方法研究写作，研究通常聚焦于童年时期的写作学习过程 [24]。这类研究采用了三种主要方法，以获取有关写作过程的不同类型信息：协议分析 [25], 、人工标注的任务驱动实验 [26], 以及击键记录 [27]。协议

分析涉及作者对写作过程进行内省，可以在写作过程中进行（同步出声思维任务），也可以在写作完成后进行（回顾性分析； [37]）。无论哪种情况，都会记录这一体验，并仔细审视其中的证据，以了解作者创作文本的潜在过程 [2]。这类方法对于识别写作所涉及的认知过程和提出假设 [38], 能够发挥关键作用，但也存在局限：它不适合检验假设，而且难以在大规模写作者样本中推广。

任务驱动实验更适合检验特定预测，并涉及对写作过程预先设定的限制或约束。这可能表现为限制或延长特定阶段可用的时间，或对每个阶段施加特定约束。虽然这种方法能够让研究者检验具体假设——例如，限制或延长初始规划阶段的时间是否会影响文本质量 [26] ——但这些约束不可避免地会干扰写作者自然进行写作任务的方式。

最后，击键记录会记录写作者输入（或删除）的每个字符，以实时考察写作。单独使用时，这种方法的优点是不会造成干扰，但通常必须与其他方法和测量相结合，包括协议分析 [39], 眼动追踪 [40], fMRI、皮肤电反应（GSR）或EEG （综述见 [27] ）。然而，若单独使用，击键记录可能会成为一种过于细粒度的测量：“通常很难将记录数据的细粒度与其背后的认知过程联系起来” [44] （第358页）。换言之，聚焦于字符级编辑，可能会使人难以将击键数据与写作的认知阶段联系起来，除非将这些数据置于其他定性或定量测量的背景下加以考察。当单独解读击键数据时，通常引起关注的是打字的缺失：例如，在其他情况下连续写作时，若击键之间的停顿超过约一秒，就会被视为聊天语境中话语层面的语言加工的表现 [45]。然而，通常不会使用这类数据来考察文本随时间在更高层级——词语、句子或段落级别——上的具体演变。换言之，单独使用击键记录时，通常用于分析写作者何时 集中精力，而不是从时间动态角度分析他们在文本中的词语或句子级别上何处 以及如何 分配注意力。

探究递归子循环

这些方法都表明，写作不仅是一项复杂的任务，本质上还是一个动态过程。作者首先进行规划：构想最终产出的粗略思路或模型 [13]。然而，这个模型必须不断调整：最终版本必须适应其组成部分的变化，例如单个句子，甚至是词语 [14,15]。例如，为了让一篇作品中的术语保持一致，哪怕只进行一次词汇变更，也可能引发更大范围的结构调整，以确保读者认为这种一致性是有意为之。其他需要调整的情况可能包括作者不满意的修辞方案，或者作者在写作过程中意外发现了对内容的新见解，或产生了新想法，从而从根本上改变最初的规划。这些调整点会促使作者反复进入规划和翻译的子循环。类比于演化博弈论中的策略研究，我们也用“探索”一词来指代规划阶段，在此期间，作者会为草稿尝试不同的选项，an

尽管人们普遍承认这些阶段彼此相关，但关于写作所涉及认知过程的研究，往往倾向于将写作的不同阶段分开测量，即便它们都纳入同一项研究中也是如此。尤其是，既有研究常将规划视为一个独立阶段，认为它始于将

词语写到纸上的行为之前，并在该行为开始前结束 [11, 26, 41–43]。本文关注修订过程中进入和退出规划与翻译阶段的转变。为了区分规划和翻译的递归子循环与涵盖规划、句子生成和修订的整体认知模型，我们将重新进入规划阶段定义为探索。换言之，写作者在修订过程中重新回到规划阶段时，会探索由自身写作所创造的、不断变化的可能性空间。当他们选择在何处进行大幅修改时，便会转回转译和转录阶段。我们通过将新的测量方法应用于一个新颖的数据集，对上述过程进行操作化；该数据集使用版本管理软件，以高于击键记录的粒度追踪修改，时间单位为分钟而非秒。

检测这些变化的能力可能会成为改进写作的重要工具的基础：随着改写点之间的间隔呈现出特定模式，我们或许能够检测出作者何时已使文本达到优化点——或者何时继续推进这一过程可能会带来递减的收益。尽管自然语言处理已在许多文本分析领域取得进展，包括通用语言理解 [16], 自动翻译 [17], 和文本生成 [18], 作者识别 [19], 主题建模 [20] 以及情感分析 [21], , 但有效写作的过程 却鲜少受到关注。据我们所知，通过分析不断演变的文本版本来研究写作的动态过程仍然十分有限，尽管已有一些研究考察了自然增长的语料库，例如推文或报纸出版物 [22, 23]。这部分是因为，这些模型受到训练所用数据的限制，无论是在数据的性质方面，还是在其通常所需的数据规模方面，都是如此。机器学习方法需要一个包含大量不同草稿的语料库，并根据其随时间的演变进行记录和系统标注。一般而言，训练数据由已完成的文本组成。记录单篇文本迭代与演变的数据——尤其是达到许多自然语言处理方法所需规模的数据——十分稀缺，甚至不存在。

在此，我们提出了一种新方法，能够在比现有自然语言处理方法小得多的数据集中，高效检测出表明写作子周期的模式；与 S-notation 等先前工作相比 [15] ，该方法的开销也小得多，因为它无需击键记录即可重构作者的活动。由于该方法依赖于许多文本编辑工具中广泛采用的自动版本管理，因此这里介绍的测量方法可用于研究多种基于文本的任务，且易于实现。这类数据使我们能够重构编辑过程的表示，以测量其时间复杂度，并指出作者何时处于探索性思维状态，轻松地将想法转化为文字，以及何时在这些任务之间快速切换。

我们采用一种新的分析方法，并使用一个新数据集。该数据集由文本版本的历时序列组成，这些文本是在前三届科学进化写作 研讨会（www.sew-workshop.org）期间撰写的，详情如下文所述。我们基于简单的编辑距离（ED），提出了三项衡量写作行为的关键指标：复杂度、探索和扭转比率。复杂度指标表明，尽管文本的设计初衷是供人们线性阅读，但写作过程从根本上说是非线性的。通过在多维空间中表示编辑事件，我们可以观察写作者探索这一空间的频率和程度，并识别写作者轻松落笔、进入“翻译心流”的时期。最后，我们可以对比探索与“翻译心流”这两个阶段，识别写作者似乎在探索与进化理论所称的“利用”之间“摇摆”的时刻。

数据收集

科学进化写作（SEW）研讨会是一系列持续开展的课程，旨在教授如何为普通读者撰写和编辑科学文本。本文报告了61名参与者的数据，他们参加了 SEW 场研讨会。这些研讨会于2019年和2020年在巴黎（索尼计算机科学实验室，N=19）、维也纳（复杂性科学中心，N=24）和伯尔尼（阿尔伯特·爱因斯坦基础物理中心，N=18）举行。参与者通过线上宣传和学术界口耳相传的方式招募。每场研讨会为期2至4天，期间参与者参加了有关写作和编辑策略的讲座。参与者均提供了书面知情同意，且年龄均在18岁以上。超过50%的参与者曾撰写过1至5篇学术作品（包括尚未发表的投稿）。研讨会的参与者包括学者（包括博士生和早期职业研究人员），其最常见的研究兴趣领域为复杂性科学、计算机科学和物理学。参与者必须提供知情同意。参与者被要求：- 将研讨会期间全部作品的使用权授予Sony CSL、萨皮恩扎大学和法国国家健康与医学研究院，但仅限于科学用途（他们保留自己作品的知识产权）；- 同意仅将其个人数据用于科学目的，且除汇总形式外不得对外披露。参与者来自不同背景，其中科学专业人士（从博士候选人到正教授）占比较高，而且英语并非他们的母语。因此，应在二语写作者这一广泛背景下评估研究结果；然而，正如范韦延等人 [30] 所示，母语写作者与二语写作者在写作过程中认知活动的出现模式相似，母语写作者的差异更大，但平均而言，其行为与二语写作者相近。此外，鉴于参与者的学术背景，他们此前已有丰富的英语写作经验，且研讨会的其他环节也以英语进行。

在这些研讨会期间，参与者接受了专业科学写作者提供的写作建议，重点是编辑策略及其与生物系统演变的相似之处。最终目标是撰写一篇至少六个段落的文章，作为一份面向大众、介绍参与者自身工作的科学报告。除此之外，这项练习没有设定其他限制，参与者的动机完全源自自身，最终目的是将最终成品用于自己的工作。研讨会期间，参与者有时间完善作品，以实践专家提供的建议。写作阶段不受监督，因此研讨会中安排的其他环节并未对工作流程产生显著影响。这项练习的开放性，以及它与参与者自身工作的紧密联系，对于研究自然写作过程至关重要，因为动机对文本质量的影响已有充分记录 [31]。

研讨会本身没有针对参与者设定排除标准，但分析仅纳入研讨会期间文本至少修改过 10 次的参与者数据，以表明其积极参与（表 1 中的数字指积极参与者）。研讨会参与者的人口构成数据并未经过系统收集。对于本研究所定义的主要指标（过程复杂度、探索和扭转比率），我们观察到三个研讨会的分布一致，表明各组之间具有广泛的一致性。详情请参见表 1。

在研讨会 1 和 2 中，参与者被要求使用 Google 文档平台撰写作品。在研讨会 3 中，参与者使用了 WeWrite 平台。这是 Sony CSL 专门开发的工具，旨在利用类网络结构化文本的导航功能，呈现并管理写作、协作写作以及管理包含多个可能替代文本的复杂项目时所涉及的复杂度。这两种工具都内置了版本管理功能，但略有不同。Google 文档会以不均匀的频率创建草稿快照，

	SEW 1	SEW 2	SEW 3
研究所	SONY CSL	复杂度 科学中心	阿尔伯特·爱因斯坦中心 基础物理学
地点	巴黎	维也纳	Bern
招募开始	2019年1月22日	2019年1月31日	2019年12月17日
招募结束	2019年1月30日	2019年2月19日	2020年1月14日
参与者	19	24	18

表1. 一张表格，详细列出了科学进化研讨会的地点、招募时间段和参与者人数。

仅在有活动的时段设置检查点。在这些时段内，检查点的频率为每分钟 ~ 1 个。相比之下，WeWrite 每 3 分钟设置一个检查点，不论是否进行写作活动。为了确保所有数据集一致，分析时排除了与前一版本完全相同的版本。

复杂性测量

要考察写作过程的结构化程度，我们必须首先定义最能代表我们希望捕捉的写作者活动的文本单位。与口语一样，书面文本可以从多个层面进行分析，例如考察字符、词语、句子、段落，甚至更大的结构单位（如较长作品中的章节）。在写作过程中，任何类型的单位都可以被删除、修改或插入。简言之，文本中任何可测量的单位都可以视为写作者行为的原子单位。不同的尺度会记录不同类型的写作活动：例如，如果我们关注的是转写（将内容编码为正字法单位的过程），那么字符级分析就能很好地捕捉这一过程。然而，如果我们更关注翻译（将想法编码为承载意义的语言单位，如词语、短语和句子的过程），字符分析可能过于细致，从而掩盖有趣的模式。例如，如果写作者修改了7个字符，仅在字符级进行分析可能无法识别出这是对一个词汇单位的修改。

适合数据集的尺度，是能够将相关（即相邻）编辑视为一个编辑事件的尺度。为确定合适的尺度，我们计算了草稿两个连续版本之间在字符、词语和段落级别的编辑距离（ED）。在每个级别，我们为每个片段标注：若有改动则标为 1，否则标为 0。这样，每个级别都会得到一个由 0 和 1 组成的序列（ $\{B\}$），用于描述该级别的编辑序列。在这个序列中，许多相邻编辑会表现为一组连续的 1。然而，例如在字符级别，许多相邻编辑很可能属于一个更高层次的事件（如更改一个词语）。因此，为了确定在哪个级别上能够最好地将相关编辑识别为单个事件，我们希望 B 中的 1 组成的簇尽可能少；换句话说，与由相同事件随机重排而成的序列相比， B 中的编辑不应更容易聚集成簇。

为对此进行测量，我们采用了 Bhattacharyya距离 [33], ，它用于衡量两个概率分布的相似度。 B 与其随机打乱版本之间距离最小的粒度级别，表明这是我们能够捕捉不同编辑事件的级别。对于我们的数据集，Bhattacharyya距离在句子级别时最小，这表明这是我们能够最佳检测原子编辑事件的分辨率。这将作为 Wagner–Fischer算法中使用的单位，该算法最终为复杂度的计算ED [32]提供元素。句子级别的编辑在我们的数据集中最为重要，这一结果与以下发现一致：主要是句子级别的翻译过程，

会调用工作记忆 [46]。

现在，设想你拿起一页文本，将其划分为若干句子，并在作者经过一段时间的修改后再次查看。其中有些句子会发生变化，有些会被删除，还有些则是全新的。就像遗传学家会做的那样，我们可以匹配保持不变的句子，并将变化视为突变。比较相邻的两个版本，我们可以设想它们代表穿过同一空间的两条路径。如果它们共享某些句子，这些句子所在的位置就是两个版本相匹配、两条路径相交之处。不匹配的位置则是作者进行干预之处。这些干预可能会创建一个全新的章节，修改作者不满意的部分，也可能影响作品的其余部分。我们该如何以一种信息丰富的方式呈现这一过程?

对于草稿的每个后续版本（第一个版本除外），我们可以定义编辑序列$\{B(t)\}$；同时要注意添加占位元素，以表示稍后将在草稿中插入的句子，从而确保随着草稿演变，未修改的句子始终对应相同的位置索引。利用这一系列序列，我们可以构建写作过程的云可视化，如图 1 的 (a) 面板所示。

借助编辑序列，我们可以探究这一过程的结构。需要说明的是，当我们在复杂度的语境下谈论结构时，指的是编辑如何映射到文本的一维空间上，即图 1(b) 的 x 轴上。如果作者在某些位置的编辑次数多于其他位置，那么编辑过程就是简单且可预测的：如果我们必须预测作者下一步会在哪里编辑，那么编辑很可能会发生在少数几个焦点位置附近。另一方面，如果编辑在文本中分布得更加均匀，那么它们就是更复杂且更难预测的过程的结果；如果我们必须预测作者下一步会在哪里编辑，那么编辑可能发生在任何位置。需要注意的是，这里我们关注的维度不是时间维度，而是空间维度：我们并非观察写作随时间推移而展开的动态过程，而是通过观察由最终成品定义的“扁平化”变化数组，获得一个表示作者在文本中将努力集中于何处的度量。编辑的空间分布代表了作者努力的分布。

我们使用香农熵 [34] 对此进行量化；如果写作付出的努力集中在空间上的某个局部，文本的熵就会较低（若反复编辑的只有一个句子，熵会趋近于 0）；而如果写作付出的努力分布在空间上的多个位置，文本的熵就会较高。为了比较大小差异显著的不同草稿，我们必须用特定规模（句子数量）作品的最大熵进行归一化。这一复杂度度量称为香农-维纳多样性指数，通常用作生态系统复杂度的替代指标 [49]。

结果

在图 1 左下方的面板中，我们可以看到将整个写作过程合成为云状图的表示。云中的每条线代表文本的一个特定版本，沿着句子的 x 轴排列，编号从 $zero$ 到各版本的句子数量上限。第一个版本显示在底部，是一条直线，从左至右依次贯穿每个句子的第一个版本。之后的版本则是锯齿线，其顶点代表句子的更新版本。句子显示得越靠上，写成的时间就越晚。最上方的锯齿线代表最后一个版本。线条采用半透明效果，以突出多个版本共有的部分。线条颜色越深，表示共享该文本部分的版本越多。

我们可以看到，作者占据了图中的大部分空间。这一证据表明，工作既不是序列性的，也不是空间局部化的。如果作者主要对文本的某个特定部分进行了修改，那么相应的版本云就会像一条水平线，只有某一部分例外。该部分会呈现出高耸于草稿其他部分之上的垂直结构。就事件序列而言，如果

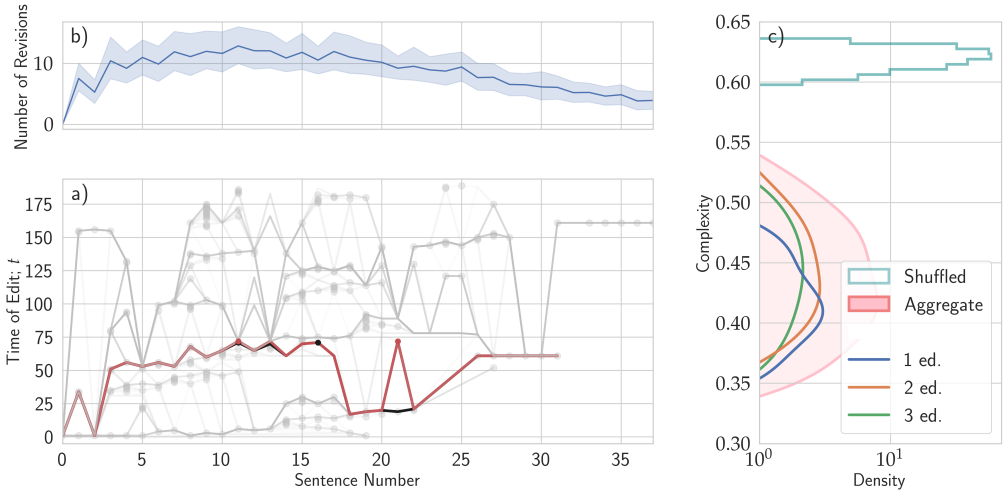

图 1. 写作过程的复杂性。(a) 普通作者生成的所有文本版本构成的云图。每个写出的句子（包括在持续写作过程中被删除的句子）都以图中的一个点表示。横轴上的位置表示某个草稿版本中句子的顺序位置，即句子编号。纵轴上的位置表示句子首次出现的草稿版本，即编辑时间 t。$\{x, t\}$ 中的一个点表示一个句子，该句子首次出现在草稿 t^{th} 版本的第 x^{th} 个句子位置。聚焦于某个特定的编辑时间，图中显示的是当时修订过的所有句子。聚焦于某个特定的句子编号，图中显示的是该句子被编辑的所有时间。每个草稿版本都以一条半透明线连接其中按顺序排列的句子来表示。不同草稿版本中保留的句子序列颜色更深。黑色 [版本 70] 和红色 [版本 71]突出显示了两个连续草稿版本的示例。第 70 版的第 21st 个句子首次出现在第 20 版中，并在第 71 版中经过编辑。(b) 我们报告所有作者每个句子的平均编辑次数。横轴显示文本中每个句子的位置，纵轴显示编辑次数。蓝色区域表示使用自助抽样程序 $[N = 1000]$计算的分布的 99.5% 置信区间。(c) 复杂度值的分布。粉色直方图显示了所有研讨会中记录的写作过程复杂度值。阴影区域为各个研讨会的核密度估计。青色线显示了通过随机置换编辑序列所生成的版本计算出的概率密度函数。

如果作者以更线性的方式处理文本，先润色文本开头，再逐步推进到后续部分，那么云图就会沿着从左下到右上的对角线扩展。这表明，虽然读者体验文本的过程是线性的，但作者创作文本的方式并非如此。

虽然编辑并非特别集中在某些位置，但也没有完全均匀地分布在文本各处。利用每次编辑的位置数据，我们可以统计每个句子经历了多少次修订。结果发现，作者在文本前半部分投入的修改更多，作品中大约第 10th 句附近的编辑次数达到峰值（图 1，左上），约位于全文进度的 30% 处。尽管这一结果应结合研讨会中所创作稿件的类型来理解（篇幅为 8–10 段的科学散文），但它有力地表明，就编辑次数而言，文本开头与结尾之间存在不对称性。

为了量化作者之间这种行为的异质性，我们统计每个句子的编辑次数，并为每位作者计算编辑分布的香农-维纳指数（SW）。若该指数为零，则表示编辑反复集中在同一个句子上。将编辑序列随机打乱后，可以得到一个分布，我们以青色显示该分布（图 1，右上图）。所有参与者的 SW 指数分布以粉色显示，并列出 SEW 每个不同版本的直方图，以突出整体重叠情况。这些结果表明，编辑过程的复杂度介于集中编辑单个句子与在全文中均匀分布编辑的随机情况之间。此外， SW 指数分布相对较广，说明不同写作者的行为存在多样性。换言之，写作者既不会将精力完全集中于一处，也不会将精力完全均匀地分散；此外，个体之间也存在差异：有些写作者的努力比其他人更为集中。

探索的测量

在写作过程中，探索——或者说在修订期间启动规划—翻译子循环——是一个难以捉摸的概念，尤其是从定量角度来看。作者探索的一些想法和修辞方案不会在文本的任何版本中留下痕迹，但许多其他想法和方案会留下痕迹，即使它们最终未能进入最终版本。为了量化探索程度，我们测量作者是否沿着从初始版本到最终版本的最直接路径推进。如果作者非常直接地沿着这条“阻力最小路径”前进，就表明文本的原始规划在整个写作过程中基本没有改变。另一方面，从初始版本到最终草稿的路径越迂回，就表明写作过程中的探索越多——换句话说，写作期间对原始规划做出的改动越多。在这项分析中，我们在计算编辑距离时分析字符级的变化。这样做是为了更细致地衡量用于探索的总体努力程度，而不是衡量变化发生在哪个语言层面。

知道草稿的最终版本后，我们可以想象许多不同的过程如何导向该版本，每个过程都可以表示为一个编辑序列。对于草稿的任意版本 t ，我们可以定义它与第一个版本（d_{0t}）及最后一个版本（d_{tf}）之间的编辑距离。再加上第一个版本与最后一个版本之间的距离（d_{0f}），便会形成一个三角形。根据定义，满足 $d_{0f} = d_{0t} + d_{tf}$ 的点集就是所有从第一个版本导向最后一个版本的最短编辑序列所在之处。换言之，当写作者或多或少沿着d_{0f} 线确定性地前进时，他们是在流畅地翻译一份预先制定的写作计划。完全沿着这条线前进的写作者，在修订阶段绝不会进入规划—翻译子循环；然而，

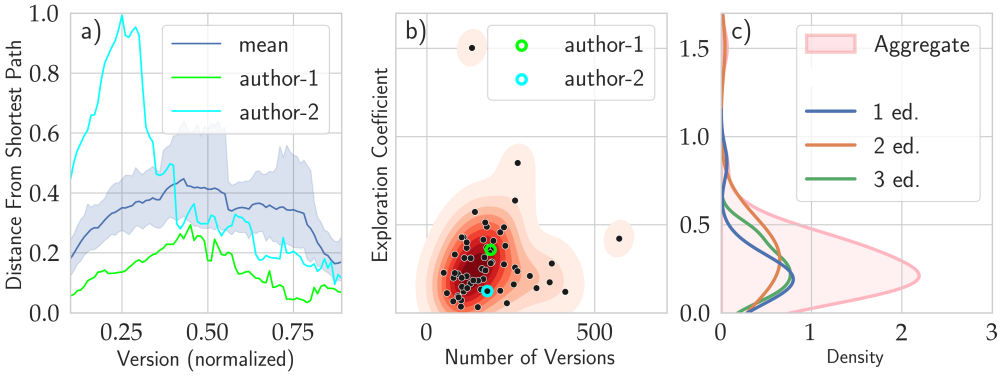

图 2。写作中的探索模式。(a) 两位作者从开始到结束的编辑阶段中偏离最短路径的距离 [以青色和绿色表示] 以及所有作者的平均行为 [深蓝色线]。最短路径由从不对文本进行修改的人实现。显然，在整个写作过程中，偏离最短路径的程度变化很大，并且越远离开头和结尾，这种偏离越明显。平均曲线以深蓝色线显示，而淡蓝色区域勾勒出使用自助法估计的 99。5% 置信区间。(b) 探索系数与每份文档的版本数量之间的关系。探索系数与版本数量之间的相关性很弱或不存在，皮尔逊相关系数=0。10，p=0。44。左侧所示两个版本的数值以相同颜色的圆圈标出。(c) 探索系数的分布，该系数定义为左侧面板中曲线下的面积，汇总于所有研讨会 [粉色]，并分别针对每个研讨会进行汇总。

由于人类写作者确实会进入这些递归子循环，我们应当预期在写作过程中的某个时刻会出现 $d_{0f} < d_{0t} + d_{tf}$。 d_{0f} 与 $d_{0t} + d_{tf}$ 之间的差异越大，写作者的探索性就越强。

结果

我们发现（图 2 左图），作者在写作过程的中间阶段与最短路径的距离达到最大。研究结果表明，在写作过程进行到一半时，平均而言，作者所写内容中有 40% 不会出现在最终版本中。这一结果意味着，作者在写作过程中会探索大量选项，例如，添加许多之后会被删去的内容。

我们引入探索系数，用以概括作者的作品中有多少内容直接促成了最终作品。我们将其定义为$H(t) = (d_{0t} + d_{tf} - d_{0f})/d_{0f}$：从若干最短序列之一出发的距离，并按首个版本与最终版本之间的距离进行归一化（以便比较不同文本）。换言之，这表示作者偏离两个版本之间最短路径的程度。已知首个版本和最终版本后，我们可以回顾性地衡量每个中间版本 t 的$H(t)$。将所有版本的贡献相加，就能衡量写作过程中发生了多少探索。由于该指标受版本数量影响，我们再除以版本总数进行归一化。因此，我们将探索系数定义为 $E = \frac{1}{t_f} \sum_t H(t)$，其中 t_f 表示版本数量；沿最短路径时， E 等于零。当然，并非所有差异都能按最广义的含义视为探索。例如，由于该指标考量字符级的编辑，细微编辑可能只是

转录事件：拼写错误、修正或细微的词汇变化。右侧面板图2c展示了所有研讨会中所有作者的探索系数直方图，并按三个研讨会分别细分。在所有作者中，我们观察到探索系数平均为28%，这意味着平均而言，作者在写作过程结束前会改动其此前已写内容的28%。

原则上，我们预计写出更多版本的作者会有更高的探索系数。中间的图2b让我们能够通过比较每位作者的探索系数与其版本数量来检验这一假设。我们观察到二者的相关性略呈正向，但并不显著（$\rho = 0.10$，p值 $= 0.44$）。这意味着，不同版本的数量并不是判断作者是否具有高度探索性的特别有效指标；拥有数百个版本的作者，可能只探索了可用空间的一小部分。

利用还是探索？扭转比率

前一种测量方式考察的是探索，也就是衡量作者在修订过程中何时返回规划阶段的指标。在这里，我们旨在测量：a）作者何时转而返回转译阶段，利用已有想法并随即将词语写到页面上；以及 b）作者何时在探索状态与转译状态之间切换。下文将介绍一种测量方法，用以反映作者在转译和转录阶段中轻松沉浸的时期，或是在探索与转译之间“扭转”的时期。

心理流畅的特征包括专注于任务、主观上感到时间飞逝，以及轻松自如的感受 [47]。写作时，我们可能会发现文本的某些部分能够轻松完成——这表明我们处于流畅状态，在此期间可以利用各种想法，并且能够毫不费力地完成令人满意的翻译和转写。文本的其他部分可能需要反复重写，这表明出现了中断，可能是因为我们对想法的翻译方式不满意，或是对想法本身不满意。心流在很大程度上通过内省来界定；在写作研究中，则是让作者报告自己的体验。与其他领域一样，写作中的心流可能具有多维性，且难以测量 [48]。虽然轻松沉浸时期可能是作者主观上正在体验更广泛的心理流畅现象的积极指标，但写作过程中的心流也可能采取其他形式，这需要进一步研究。因此，我们特指“翻译流”为句子生成阶段中的一种轻松沉浸状态。

为了量化这一点，我们考察连续版本三元组，例如版本 A、B和 C。假设 ED 与 A 之间的距离等于 A 与 B 之间的距离和 B 与 C 之间的距离之和（$AC = AB + BC$）。在这种情况下，当写作C时，作者没有覆盖其在写作 B时所做的修改，这表明作者处于翻译流状态。相反，如果在 C 中，作者覆盖了在 B中所做的修改，那么他们对自己的作品并不满意。在这种情况下，我们说作者正在转回规划与探索，无法在翻译和转写中保持势头。如果作者能够流畅且不间断地生成令人满意的文本，那么该空间中的轨迹将是一条从第一个版本到最后一个版本的直线。偏离这条直线意味着作者相对于前一版本做出了一些修改，但这些修改可能并不令人满意，也可能很快被放弃。

借助扭转比率，我们将探索与翻译流之间拐点的相对数量，测量为文本连续版本之间欧几里得空间中的角度。由于编辑距离是一种定义良好的度量，因此这一过程是可行的。不过，我们必须确定使用多少个维度。高维空间可以确保保留各版本之间的所有距离，

但会丢失大量结构。通过主成分分析，我们确定，三维空间能够保留 90% 以上的信息。

使用 t 分布随机邻居嵌入（t-SNE），我们可以定义一个三维嵌入，在其中距离得以保留，并更加重视短距离的保留，从而保留邻近点的几何结构。因此，我们可以在欧几里得空间中定义每段文本从初稿到最终形式所经历的完整轨迹。图 3 展示了文本在欧几里得空间中的轨迹示例（以及一个三元组的轨迹）。我们仅以二维展示轨迹，因为肉眼可辨的大多数几何属性在二维中已经显现。箭头表示从一个版本到下一个版本的每一步。坐标轴上的绝对值表示各草稿在嵌入空间中的位置，因此其具体数值并不重要。两点之间的差异表示两个版本之间的字符 ED ，即一个版本变成另一个版本所需编辑的字符数。

在欧几里得空间中，这三个距离构成一个角度 β，位于 B中，如图 3 的插图所示。在翻译流的情况下，三个版本会落在一条直线上并$\beta = 180°$；而在探索的情况下 $-180° < \beta < 180°$。为了更好地量化探索与翻译流之间的差异，我们任意将所有角度 β 处于区间 $[30°, 150°]$ 及其相反区间$[-30°, -150°]$内的事件定义为探索。相应地，我们将角度处于区间 $[-30°, 30°]$ 和 $[-150°, 150°]$内的事件定义为翻译流。这一划分如图 3（插图）所示，其中展示了 A 和 C （实心圆）的固定位置，以及B的假设版本。如果作者位于绿色阴影区域，则版本 B 与版本A 和 C 足够接近，因此我们可以认为作者处于翻译流状态，沿用已有计划。然而，如果 B 远离 A 和 C ——即蓝色区域——则将其视为探索。因此，扭转比率定义为作者处于翻译流状态的版本所占比例。

结果

在图 4 的左侧部分，我们报告了所有作者的所有轨迹中观察到的角度分布。我们发现，作者平均有 1% 到 2% 的时间处于探索状态，这由图 4a 中灰色阴影区域表示。图 4a 中的深蓝色线表示作者之间平均 β 角度的分布（显示为最短路径 $180°$ 与 β 之间的距离），大多数作者约有 1.6% 的时间处于探索状态。换言之，作者大部分时间都处于翻译流中。图 4a 展示了两位特定作者的数值，表明有些作者从未进入完全探索状态：例如，作者1 在草稿之间沿“阻力最小路径”行进时，偏离幅度始终只有约 10-20° 。与其他测量指标一样，这显示出显著的个体差异；然而，图 4b 表明，这种情况在不同研讨会之间大致保持一致。

最后，达到最终成品所进行的编辑总次数越多，扭转比率就越呈现系统性下降（$\rho = -0.46$ 且 p值为 < 0.001）。换言之，编辑较少的写作者似乎会进行更多探索。图 4c显示，总体编辑次数较少的文本更可能出现较高的扭转比率。该图还标出了图 4 中的写作者；作者1进行的探索性尝试所涉及的编辑事件少于作者2。换言之，作者2可能进行了更多编辑，但作者1探索得更多。

讨论与结论

本文提出了一套依托常见版本控制系统分析写作过程的新方法。这些方法能够为我们提供详细信息

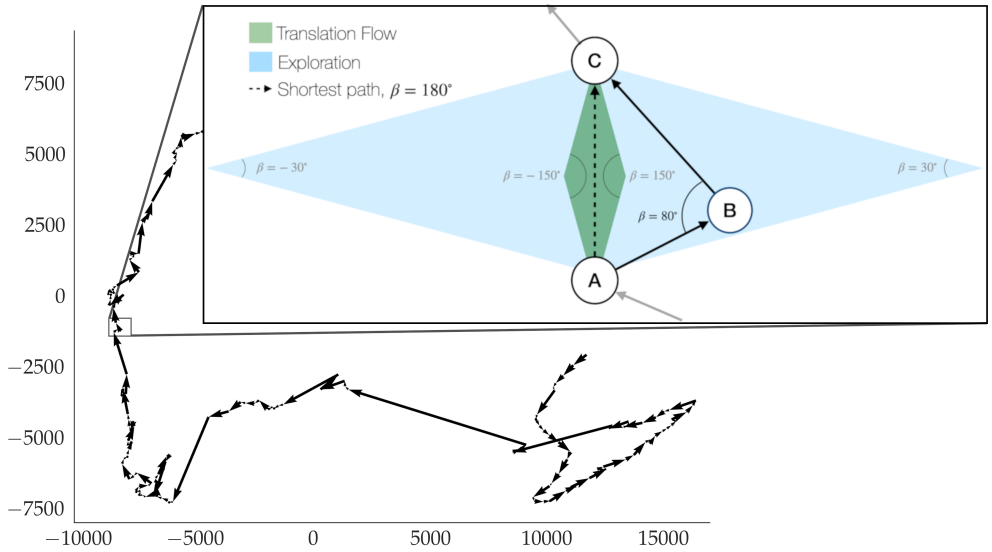

图3. 草稿轨迹。展示单个作者的草稿版本在二维空间中的 tSNE 嵌入示意图。该空间中的距离表示一个草稿与另一个草稿之间编辑的字符数。箭头描绘了单个草稿的创作轨迹。右上方的示意图展示了轨迹中的一个示例片段，以说明计算扭转比率时所用的向量，以及探索与翻译心流之间的阈值。

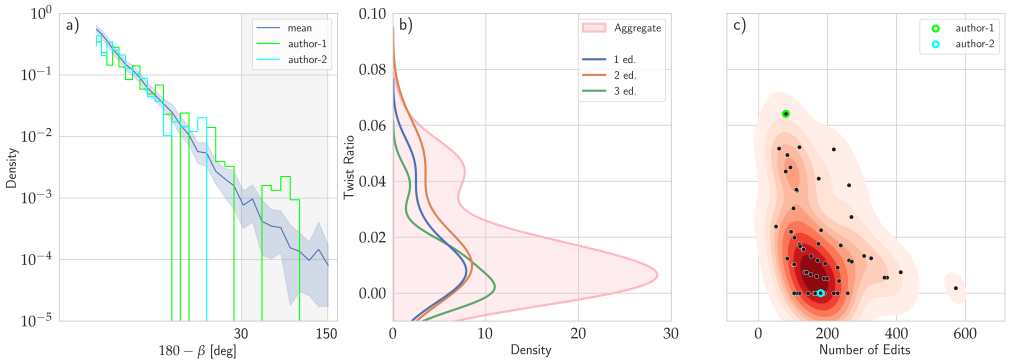

图4. 扭转比率(a) 草稿演变过程中连续速度之间的角度分布。我们定义为探索性的区域以灰色突出显示。深蓝色线表示角度的平均分布，淡蓝色区域表示使用自助法在各点估算出的 99.5% 置信区间。浅绿色线和青色线展示了两个示例版本。(b) 所有研讨会 [pink] 中的扭转比率分布，以及各研讨会的核密度估计。(c) 每份草稿的探索步骤分布，并与各研讨会的编辑次数进行比较。左侧面板中的示例以相应颜色勾勒。

关于作者的活动，例如他们在何处、何时工作，以及他们在多大程度上探索新的路径，而非遵循预先规划的路线。总体而言，这让我们得以了解整个过程有多随机、机械或复杂。我们提出的新指标对所有这些方面进行了定量描述，并通过对写作活动快照进行简单的编辑距离度量，为写作和修订过程提供了关键见解。

复杂度测量提供了一种观察写作者在文本中付出的努力如何进行空间分布的方式。观察到的分布表明，在我们的特定写作任务中，努力程度介于两个极端之间。写作者往往不会集中精力处理文本中的某一部分（这会使复杂度分数为零），但他们的努力也并非随机分布于文本之中。归一化复杂度分数几乎处于这两个极端的中点，这表明写作者在整篇文本中会间歇性地投入精力，并在约 30% 处出现一个较低的峰值。此外，各作者的复杂度分数分布并不均一：没有写作者会将精力完全集中于某处，也没有人会将其完全分散；但有些人比其他人更专注，而另一些人则表现出更分散的努力分布。

我们的数据还表明，尽管对读者而言，文本是序列性的实体，但作者并非以这种方式创作文本。文本的不同部分不仅投入的精力各不相同，而且这些精力的分配也并非序列性的——作者并不是从引言一路写到结论。未来研究可以运用我们的测量方法，探究哪些因素可能影响这种个体差异，以及这种差异可能如何与成稿的特性相关。例如，熟练程度较高的作者或许会更均匀地将精力分配到文本各处，而经验较少的作者一开始可能会在引言段落上过于纠结。我们在此介绍的测量方法还可以与协议分析等更具质性的方法结合，以探究写作障碍等现象的体验是否与文本中精力的分布相对应：我们在感觉写作吃力的部分上投入的精力会更多还是更少？当我们试图以序列性的方式处理一个本质上更适合分散进行的过程时，是否会更感吃力？

转向修订过程，我们引入了相关指标，用于分别衡量写作者在规划与探索之间，以及翻译与转写之间转换的潜在迹象。探索系数的分布证实，写作并不是一个严格按照从规划到句子生成再到修订这一顺序进行的线性过程；有效的修订包含探索与翻译的递归循环。如果这一过程是线性的，我们应当观察到更低的系数值，这表明作者坚持预先设定的计划，仅在草稿之间的最短路径上做出极少偏离。然而，我们观察到的恰恰相反。与最短路径的平均距离在整个过程的中段附近达到最高值（40%），直到工作的末尾（版本历史的 85 – 90%）才明显下降。这一结果表明，在最终修订事件发生之前，几乎一半的文本都必须改变，才能形成最终成品。观察到的行为与这样一种模型并不相容：在该模型中，规划文本和写出词语是两个彼此割裂的活动。尽管早期研究已经认识到这一点，但对其进行测量一直较为困难；我们的探索系数为检测并量化文本在修订过程中对整体规划所做的调整提供了一种方法。

个体的探索曲线呈现出多种多样的曲线形态，且波动性很高。波动性表明，写作过程中经常会发生重组事件。不过，这些事件发生的时间可能为其背后的认知活动提供不同的解释 [36]。例如，活动开始时进行修订，可能表明难以着手；而在活动结束时进行修订，则可能有不同目的，例如提升连贯性。值得注意的是，对于作者作品中转折点的分布，我们发现其呈重尾分布，

剧烈变化出现的频率越来越低。扭转比率随着编辑次数的增加而下降，这表明随着草稿投入的工作越来越多，探索会变得更加困难，这可能是因为限制条件增多。复杂度测量表明，投入的精力并未按照读者阅读文本时所经历的顺序分配，也没有均匀分布在文本各处。从作者推进最终版本的时间动态中也能看出这一点：每次编辑未必都会让文本以相同幅度接近最终版本，但随着过程接近尾声，作者会更快地推进至最终版本。

尽管我们的结果揭示了一种观察写作各阶段的新颖视角，但当然也存在一些局限。首先，体裁可能对我们观察到的模式产生了很大影响。简短的科学文本需要准确地进行框架构建，以便最好地呈现正文中的核心结论；这或许可以解释为什么我们的作者将更多精力放在文本开头。也可以将这一效应归因于作者对读者注意力的判断：他们认为，文本开头需要更加精确，才能激发读者的兴趣。研讨会的时长以及文本长度的大致范围，也可能影响了结果。这些动态在更长时间的写作过程和更长的文本中会如何表现，仍有待观察；例如，在我们样本的文本中，大约第10句话附近出现了编辑高峰，约位于全文的30%处，这可能是文本长度造成的假象。这或许反映了作者对文本前三分之一部分的关注，也可能反映了他们更集中地关注引言段落。我们可以预期，诗歌或歌曲创作等其他体裁的写作会呈现截然不同的动态。尽管如此，这些测量指标为我们提供了新的方法，用于检验有关修订如何因体裁、经验水平和文本篇幅而异的预测。

最后，我们讨论的方法也适用于合作作品，而这类作品可能呈现出与单人创作截然不同的动态。事实上，科学出版物尤其常由多位作者合作撰写。我们在此提出的度量可应用于由多位作者共同创作的单篇文本，从而揭示不同作者在同一文本中各自将精力集中于何处。作为作者，在合作撰写中面临的一项复杂挑战，是在他人同时创作同一文本的情况下，理解何时以及何处应集中精力。例如，了解作者A的扭转比率变化规律，可能会让作者B意识到，复杂文本中的某个部分即将完成，可以进行审阅了。

未来几年，新的人工智能驱动的自然语言处理技术将深刻影响写作实践。现有的写作助手能够提供修正和替代表述，却很少能改善作者自身的写作过程。我们提出的指标为这一过程提供了创新的描述，可应用于写作工具。这些工具能够建议作者关注文本中被忽略的部分，指出作者何时正处于高效的翻译流中，或何时正在“扭转”，此时或许适合休息一下。换言之，这些围绕对人类写作认知过程的深入理解而设计的工具，有助于作者充分发挥潜能。

致谢

作者谨向 Vittoria Colizza 致以诚挚谢意，SEW 项目正是与她共同构思并启动的。作者还要衷心感谢 Mark Buchanan 和 Justin Mullins 在 SEW 研讨会期间作出的宝贵贡献，以及一系列源源不断、富有启发性的交流。最后，谨向 Elisabetta Falivene（罗马智慧大学）和 Riccardo Corradi（iLab）致以诚挚谢意，感谢他们在组织 SEW 研讨会方面作出的宝贵技术和组织贡献。

以下机构为 SEW 研讨会的组织工作提供了支持：Sony CSL Paris、法国国家健康与医学研究院（Institut National de la Sant´e et de la Recherche M´edicale）、罗马智慧大学、维也纳复杂性科学中心（Stefan Thurner 教授）、伯尔尼阿尔伯特·爱因斯坦基础物理中心（Gilberto Colangelo 教授）、巴黎复杂系统研究所（Institut des Syst`emes Complexes，David Chavalarias 教授）。作者还要感谢 CREF-SONY 联合倡议在本作品编辑阶段提供的支持。

参考文献

1. Kankaraˇš M, Montt G, Paccagnella M, Quintini G, Thorn W.《技能至关重要：成人技能调查的进一步结果》。经合组织技能研究。经合组织出版社。2016；。2. Hayes JR, Flower LS. 写作研究与作者。《美国心理学家》。1986；41(10):1106。3. Sinharay S, Zhang M, Deane P. 使用数据挖掘方法，根据写作过程和成品特征预测文章得分。《教育应用测量》。2019；32(2):116–137。4. Leijten M, Van Waes L, Schrijver I, Bernolet S, Vangehuchten L. 描绘硕士生在第一语言和第二语言基于来源的写作中对外部来源的使用情况。《第二语言习得研究》。2019；41(3):555–582。5. Breetvelt I, Van den Bergh H, Rijlaarsdam G. 写作过程与文本质量之间的关系：何时以及如何？《认知与教学》。1994；12(2):103–123。6. Baaijen VM, Galbraith D, de Glopper K. 写作信念和规划对写作表现的影响。《学习与教学》。2014；33:81–91。7. Baaijen VM, Galbraith D. 通过写作进行发现：与写作过程和文本质量的关系。《认知与教学》。2018；36(3):199–223。8. Lardilleux A, Fleury S, Cislaru G. Allongos：用于写作者草稿生成研究的纵向对齐。载于：国际智能文本处理与计算语言学会议。施普林格；2013。第537–548页。9. Cotos E, Huffman S, Link S. 了解研究生写作者在修订期间与研究写作辅导工具的互动及其影响。《写作研究杂志》。2020；12(1)。10. L´opez P, Torrance M, Fidalgo R. 在线管理写作过程及其对小学高年级学生文本质量的贡献。《心理主题：心理学杂志》。2019；31(3):311–318。11. Limpo T, Alves RA. 规划策略对写作动态和最终文本的影响。《心理学学报》。2018；188:97–109。12. Arrimada M, Torrance M, Fidalgo R. 向一年级写作者教授规划策略的效果。《英国教育心理学杂志》。2019；89(4):670–688。13. Pickering MJ, Garrod S. 迈向对话的机制心理学。《行为与脑科学》。2004；27(2):169–190。

14. Gell-Mann M.《夸克与美洲豹：简单与复杂的历险》。麦克米伦；1995。15. Eklundh KS, Kollberg P. 话语结构的形成：计算机辅助篇章分析如何揭示大学生写作中的整体修订。《语用学杂志》。2003；35(6):869–891。16. Devlin J, Chang MW, Lee K, Toutanova K. BERT：用于语言理解的深度双向变换器预训练。arXiv预印本 arXiv:181004805。2018；。17. Kumar J, Henge SK, Dubey P. 神经网络方法机器翻译的深入文献综述。《计算领域近期创新》。2022；第45–59页。18. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. 语言模型是少样本学习者。arXiv预印本 arXiv:200514165。2020；。19. Fabien M, Villatoro-Tello E, Motlicek P, Parida S. BertAA：用于作者归属判定的BERT微调。载于：第17届国际自然语言处理会议（ICON）论文集；2020。第127–137页。20. Vayansky I, Kumar SA. 主题建模方法综述。《信息系统》。2020；94:101582。21. Mohammad SM. 情感分析：自动检测文本的效价、情绪及其他情感状态。载于：《情绪测量》。爱思唯尔；2021。第323–379页。22. Cui W, Liu S, Tan L, Shi C, Song Y, Gao Z, et al. Textflow：旨在更好地理解文本中演变主题。《IEEE可视化与计算机图形学汇刊》。2011；17(12):2412–2421。23. Shi F, Teplitskiy M, Duede E, Evans JA. 极化群体的智慧。《自然·人类行为》。2019；3(4):329–336。24. Allal L, Chanquoy L, Largy P.《修订：认知与教学过程》……第8卷。施普林格；2004。25. Emig J. 十二年级学生的写作过程；1971。26. Limpo T, Alves RA, Fidalgo R. 儿童的高阶写作技能：规划与修订能力的发展及其对写作质量的贡献。《英国教育心理学杂志》。2014；84(2):177–193。27. Lindgren E, Sullivan K. 观察写作：击键记录和手写带来的启示。博睿；2019。28. Leblay C, Caporossi G. 在线写作数据可视化的图论方法。《十字路口》。2015；第171页。29. Levenshtein VI. 能够纠正删除、插入和逆转的二进制码。载于：《苏联物理学报告》。第10卷；1966。第707–710页。30. Van Weijen D, Van den Bergh H, Rijlaarsdam G, Sanders T. 第二语言写作中的第一语言使用：一项复杂现象的实证研究。《第二语言写作杂志》。2009；18(4):235–250。31. Bandura A, Free man W, Lightsey R. 自我效能：《控制力的运用》；1999。

32. Wagner RA, Fischer MJ. 字符串到字符串校正问题。《美国计算机学会会刊》。1974;21(1):168–173. 33. Bhattacharyya A. 关于由概率分布定义的两个统计总体之间散度的一种度量。《加尔各答数学学会通报》。1943;35:99–109. 34. Shannon CE. 通信的数学理论。《贝尔系统技术期刊》。1948;27(3):379–423. 35. Hayes J R. 写作的建模与再建模。《书面交流》。2012;29(3):369–388. 36. Bergh Hvd, Rijlaarsdam G. 写作过程中观点生成的动态。《写作研究》。2007 Jan 1;20:125. 37. Smagorinsky, Peter. 协议分析的信度与效度。《书面交流》1989, 6(4):463–479. 38. Nisbett RE, Wilson TD. 述说超出我们所能知晓的内容：关于心理过程的言语报告。《心理学评论》。1977 May;84(3):231. 39. Perrin D. 进程分析（PA）：研究工作场所中的写作策略。《语用学杂志》。2003 Jun 1;35(6):907-21 40. Wengelin ̊A, Torrance M, Holmqvist K, Simpson S, Galbraith D, Johansson V, Johansson R. 结合眼动追踪与击键记录的方法：研究文本生成中的认知过程。《行为研究方法》。2009 May;41(2):337-51. 41. Evmenova AS, Regan K, Boykin A, Good K, Hughes M, MacVittie N, Sacco D, Ahn SY, Chirinos D. 使用计算机图形组织器，强调文章写作中的规划。《特殊儿童》。2016 Jan;82(2):170-91. 42. Llaurado A, Dockrell JE. 儿童的写作计划：特点及其对写作表现的影响。《读写能力研究杂志》。2019 Sep;51(3):336-56. 43. Rostamian M, Fazilatfar AM, Jabbari AA. 规划时间对认知过程、监控行为及第二语言写作质量的影响。《语言教学研究》。2018 Jul;22(4):418-38. 44. Leijten M, Van Waes L. 写作研究中的击键记录：使用 Inputlog 分析并可视化写作过程。《书面交流》。2013 Jul;30(3):358-92. 45. Chukharev-Hudilainen, E. 自发书面交流中的停顿：一项击键记录研究。《写作研究杂志》。2014;6(1):61-84. 46. Kellogg RT. 书面句子生成中的工作记忆成分。《美国心理学杂志》。2004 Oct 1:341-61. 47. Csikszentmihalyi M. 超越厌倦与焦虑。Jossey-bass; 2000.（原作发表于1975年）。48. Ainley M, Enger L, Kennedy G. 难以捉摸的“心流”体验：定性与定量指标。《国际教育研究杂志》。2008 Jan 1;47(2):109-21. 49. Keylock CJ. Simpson多样性和Shannon–Wiener指数是广义熵的特例。《Oikos》。2005 Apr;109(1):203-7.