

基于语料的新诗技：机器诗歌美学探源

李 睿

内容提要 近年来，伴随着互联网和人工智能的发展，出现了文学创作的自动化浪潮。大型科技公司和高校纷纷在智能写作机器人领域加大了研发投入力度，诞生了诸如“小冰”“偶得”“薇薇”等作诗机器人。2017年，文学史上第一部百分之百由“机器诗人”创作的汉语诗集《阳光失了玻璃窗》出版发行。为了剖析机器诗歌与人类诗歌的差异，本文回溯到机器写作历史的开端——德国斯图加特诗派的机器诗歌实验，尝试探索机器诗歌的美学源头——本泽的信息理论美学。本泽通过对“自然诗”与“人工诗”的区分，确定了机器诗歌的美学特征，为文学艺术与信息技术的结合奠定了美学基础。

关键词 人工智能 机器写作 机器诗歌 信息理论美学

当代人工智能机器文学创作的划时代事件发生在2017年5月19日——由微软亚洲研究院研发的人工智能作诗软件“微软小冰”的诗集《阳光失了玻璃窗》出版发行，成为文学史上第一部百分之百由“机器诗人”创作的诗集。微软小冰并不是国内唯一的写作机器人，近些年研发的智能机器诗人除了会写现代诗的“小冰”外，还有IBM公司研发的作诗机器人“偶得”以及清华大学研究团队开发的“薇薇”和“九歌”等。“机器诗人”如雨后春笋般涌现，读者自然会拿机器诗人的作品与人类诗人的作品相比较。遗憾的是，无论是现代文学学者还是当代诗人都对新诗机器人“小冰”的创作“才能”表示了一定程度的否定。学者认为小冰的诗没有情感、想象力以及经验构造能力，诗人则指出小冰的诗缺少灵性，只是一种“词语的游戏”。^①然而，他们指出的机器诗歌的这些“弱点”，却恰恰是机器诗人的创作特点，也是机器诗歌区别于人类诗歌的根本所在。为了进一步剖析机器诗歌与人类诗歌的差异，我们有必要回溯到机器写作历史的开端，去探索机器诗歌的美学源头——信息理论美学（Informationstheoretische Ästhetik），考察机器写作的美学特征。尽管创作主体发生了变化，但机器文学的审美功能因其特有的美学特征而得以保留，相较人类作家的作

^① 宋俊宏《人工智能与文学创作》，载《长江文艺评论》2017年第5期，第9页。

品，其审美功能甚至还会得到加强，因为机器诗歌只能通过美学信息来发挥审美功能。

一、机器诗歌初试——卢茨的“随机文本”

如今，学界研究人工智能机器写作历史，大多会回溯到1962年美国诗歌创作软件“自动垮掉的一代”（Auto-Beatnik），其代表作是发表在《地平线》杂志上的《玫瑰》《孩子们》《姑娘》《风筝》《牛排》等诗作。事实上，人类历史上第一组机器创作的诗歌是德语诗歌，是1959年由斯图加特工业大学的计算机科学家卢茨（Theo Lutz，1932—2010）使用楚泽Z22（Zuse Z22）大型计算机开发的机器作诗程序“随机文本”（Stochastische Texte）“写”出来的。所谓“随机文本”是指“其语法结构是预先给定的，但其词汇是随机生成的文本”^①。卢茨为这个程序选择的语料词典是德语作家卡夫卡的小说《城堡》中的词汇，卢茨将这些语料导入数据库，在楚泽Z22上运行，并使用指定的概率矩阵，算法便会将它们随机组合为诗句，这个文本生成器是机器文学创作领域的最早实践。

在二十世纪五十年代末六十年代初，大型电子计算机刚刚发展到第二代，那时还没有台式机，计算机系统也还没有图形用户界面，只有监视器，监视器的屏幕仅能以代码形式来显示数字。当时大型计算系统几乎专门用于运行复杂的数学计算，尤其是随机运算，因此，将计算机运算过程应用于文学文本的生成在当时极具开创性意义，这也使得科学家们开始进一步探索机器书写的可能性，极大鼓励了机器文本创作的实验。卢茨就是在这样的科技史背景下进行机器诗歌实验的。他为楚泽Z22计算机系统开发了一个程序，该程序使用随机生成器，可以基于对主题、谓词、逻辑运算符、逻辑常数和单词“是”的特定选择来生成文本。计算机系统进行选择语料的基础是卡夫卡小说《城堡》中的词汇，计算机生成的结果最后被打印出来，这就是文学史上的第一组机器诗歌，比如下面这组诗句：

NICHT JEDER BLICK IST NAH. KEIN DORF IST SPAET.
EIN SCHLOSS IST FREI UND JEDER BAUER IST FERN.
JEDER FREMDE IST FERN. EIN TAG IST SPAET.

^① Theo Lutz, „Über ein Programm zur Erzeugung stochastisch-logistischer Texte“, in *Grundlagenstudien aus Kybernetik und Geisteswissenschaft*, Jahrgang 1, Nummer 1 (1960), S. 12.

JEDES HAUS IST DUNKEL. EIN AUGEN IST TIEF.
 NICHT JEDES SCHLOSS IST ALT. JEDER TAG IST ALT.
 NICHT JEDER GAST IST WUETEND. EINE KIRCHE IST SCHMAL.
 KEIN HAUS IST OFFEN UND NICHT JEDE KIRCHE IST STILL.^①

并非每一瞥都是近的。没有村庄是晚的。
 一座城堡在野外而每个农夫都是远的。
 每位陌生人都是远的。一日是晚的。
 每幢房子都是暗的。一只眼是深的。
 并非每座城堡都是老的。日日皆是旧的。
 并非每位宾客都是怒的。一座教堂是瘦的。
 没有房屋敞开着而并非每座教堂都是寂静的。

卢茨的机器诗歌创作是诗歌艺术与计算机技术之间的第一次交叉实验，他尝试将概率运算、随机运算与文学实验相结合，发现诗歌文本在计算机的控制之下可以脱离作者的主观性。尽管也有文学批评家从传统诗歌理论出发，认为卢茨这样的计算机科学家的机器诗歌实验并非是以严肃态度进行的文学创作，某种意义上讲，这种“创作”更像是在用计算机这种功能强大的实验设备来完成一次与语言的“嬉戏”。^②

二、机器诗歌溯源——斯图加特诗派

卢茨研发机器作诗程序“随机文本”并非一次纯个人实验，他是斯图加特大学哲学教授、信息理论美学奠基人马克斯·本泽（Max Bense, 1910—1990）的学生。在二十世纪五六十年代，围绕着本泽，形成了世界上第一个机器诗歌团体——“斯图加特诗派”（die Stuttgarter Gruppe/Schule），该诗派由一群边进行智能机器写作实验边探索美学理论创新的诗人组成，卢茨即其中一员。可以说，从一开始，德国的机器诗歌创作就是与信息理论美学同时发展的，以本泽为中心的斯图加特诗派坚持将作诗程序研发、机器诗歌创作与美学理论创建这三种不同的“实验”同时进行。

说起机器诗歌的起源，特别是论及“机器诗歌”这一概念的形成及其追求的诗

① Theo Lutz, „Stochastische Texte“, in *Augenblick. Zeitschrift für Tendenz und Experiment*, Okt./Nov. (1959), S. 9.

② Gerhard Stickel, „Monte-Carlo-Texte Automatische Manipulation von sprachlichen Einheiten“, hrsg. von William E. Simmat, *Kunst aus dem Computer*, Stuttgart: Verlag Nadolski, 1967, S. 54.

学目标，我们还要回到产生“随机文本”的那个时代——二十世纪五六十年代，当时在欧洲正是实验诗歌如火如荼急于寻找全新的诗意表达方式之时。欧洲的新先锋派诗人继承了达达主义、未来主义、立体主义和超现实主义诗人的实验传统，主张通过思想交流来彼此影响，他们致力于寻找新的诗歌材料、表达方式与写作技巧，并将发表形式与其他艺术形式相结合，他们中的激进者坚持认为“最完美的诗人恰恰是最无个性的人”。^①以偶然性为美学原则的“随机写作”获得新先锋派诗人的青睐，从此开启了文学艺术与科学技术的互动发展历程。与这些发展并行的是接受者的角色也被重新定义了，接受者从被动的欣赏者转变为可以进行积极互动的创作参与者。这正是机器诗歌产生的文学史背景。可以说，从一开始，机器诗歌创作的最初动因就是文学性的。新先锋派诗人的文学实验在那个时代恰逢计算机走入人类生活，诗人们在寻找全新诗意表达方式的过程中，将眼光投向了信息和数据处理领域，也就是当时刚发展到第二代的大型电子计算机上。于是，诗人的创作冲动与计算机技术相结合产生了第一代机器诗歌，这从某种意义上讲也实现了一种“诗歌乌托邦”——即诗歌终于可以被源源不断地、“无限地创作”出来了。诗人们当时应该还没有意识到，从那时开始，作为生成和存储媒介的计算机将要改变文学的创作方式和接受方式，成为实验文学的重要工具。这一时期的诗人与数学、信息论、控制论、博弈论以及符号学有着密切联系。在这些实验诗人团体中，最著名的尝试者就是斯图加特诗派。

“斯图加特诗派”这个概念最初是本泽在1961年的一次讨论会上提出来的，用来指称当时在德国斯图加特围绕着本泽发展起来的实验诗歌团体。斯图加特诗派借鉴了达达主义、立体主义以及控制论和语言哲学的艺术实验传统。斯图加特诗派的诗人们倾向于跨界，因此形成了一个跨越艺术、学科与文化的诗歌交流平台。斯图加特诗派意图通过机器诗歌来变革诗歌概念，颠覆传统诗歌观。他们认为，传统意义上的诗歌是用来表达主观情感的浪漫主义话语，而斯图加特诗派则要创作出一种以语料为基础的、数字化的、取消主观性的诗歌，以此“诗技”来完成一种美学的进步和文学的进步。在《斯图加特诗派宣言》中，本泽这样说道：

审美创造不再依赖于煽情或那种木乃伊化的神秘创作意图而进行；[……]
实现审美构建需要作者、印刷者、画家、音乐家、翻译者、技术人员和程序员。
我们要谈的是“语料诗或[语料]艺术”。以诗匠取代[传统意义上的]诗人—先

^① Egon Friedell, *Ecce Poeta*, Berlin: S. Fischer Verlag, 1912, SS. 212–213.

知以及情绪的小丑，诗匠操控语料并使语料处于并保持被处理的和持续运转的状态。[……]这才是“诗技”[Poietike techne]，是一种进步的美学或诗学，这种有意识的[技术]应用证明了文学的进步，就像科学的进步一样。^①

本泽认为，传统意义上的诗人要么是“诗人先知”，要么就是“情绪的小丑”，作诗是一种依赖于主观创作意图的非理性情绪表达的过程，传统诗歌的特征就是主观性；而斯图加特诗派的诗歌创作则要以语料为基础，让“诗匠”去操纵那被“喂”满语料的机器，来创作出一种不再依赖于主观性的“语料诗”。语料诗也就是机器诗歌，这是一种数字化的、去主体化的“技术诗”（Technologische Poesie）。

斯图加特诗派依托德国斯图加特工业技术大学（即斯图加特大学），利用校计算机中心配备的当时世界上最先进的大型计算机系统吸引了大批诗人、艺术家和科学家来尝试这项新“诗技”，他们使用计算机程序进行诗歌生成实验，成功创作了人类文学史上的第一代机器诗歌。这些由计算机生成的“随机文本”在形式上都很相似：它们排列成行并形成诗节；缺少节奏和音步，也鲜有押韵形式，主要都是自由诗体；文本的内容往往还会引起争议。而微软小冰创作的新诗也具有这些形式特征。机器诗歌显然与传统诗歌相去甚远，但“却与某些现代诗惊人相近”^②。作诗机器与人类诗人差异更大：人类诗人具有主观创作意图，作诗时会有意识地选词造句、字斟句酌，直到满意为止；而计算机并没有这种主观选择能力。^③斯图加特诗派诗人认为，如果将来信息技术能够实现将选择功能内置到程序中，并且计算机可以运行所有选项，直到创作出最有诗意的版本为止，那么终有一天，机器诗人将能比肩人类诗人，就此而言，作诗机器的研发也是一个无限开发人类创造潜力的过程。斯图加特诗派诗人们的兴趣当然不止于研发作诗机器，他们只是将设计程序视为“写诗”这一文学创作过程中一个必经步骤，他们更大的兴趣在于这种程序所“写”出来的机器诗歌。卢茨曾指出：“这些[作诗]程序的存在清楚地表明，由程序控制的电子计算机系统的使用实际上并不局限于与数字概念相关的问题。这样的程序给‘计算’这一概念赋予了更一般的含义……机器工作本身并不重要。唯一重要的是如何解释机器的[作诗]功

① Max Bense und Reinhard Döhl, Zur Lage. Manifest der Stuttgarter Gruppe, 1964, http://www.stuttgarter-schule.de/zur_lage.htm [2019-05-27]

② Manfred Krause/Götz F. Schaudt, *Computer-Lyrik. Poesie aus dem Elektronenrechner*, 2. erweiterte Auflage, Düsseldorf: Droste Verlag, 1969, S. 15.

③ Samuel R. Levin, “On Automatic Production of Poetic Sequences”, in *Texas Studies in Literature and Language. A Journal of the Humanities*, No. 1 (1963), S. 146.

能……我们的[作诗]程序的任务就是要进行复杂的随机文本的创作。”^①

机器诗歌作为一种文学体裁，首先必须满足文学标准，具有文学性。考察机器诗歌文本，我们发现，从第一代机器诗歌——卢茨的“随机文本”（1959）开始，到微软小冰的诗集《阳光失了玻璃窗》（2017），机器诗歌都具有诸如大胆的联想和跳跃、通过两个或多个意象碰撞而产生想象图景、独特的诗意色彩等文学特征。这些特征在人类诗人的作品中也会有，它们在机器诗歌中甚至得到了加强。之所以如此，与机器诗歌的美学特征有关。本泽总结过斯图加特诗派机器文学实验的目标：“数字文本的语言游戏旨在告知外部世界语义的丧失，以赢得审美的收获。”^②本泽直面文学与技术，尝试着为二者的结合奠定美学基础，为机器文学确定了独特的美学特征。

三、机器诗歌美学——本泽的信息诗论

根据斯图加特诗派进行的机器诗歌实验，本泽发展出信息理论美学，这种美学被认为是一种“客观的、在自然科学意义上的精确美学”^③。哲学家本泽自己也是诗人^④，他的目的是要发展出一套能适应计算机科技发展水平的美学语言，该语言的词汇库必须包含并汇集文学、哲学、文艺理论、物理学、信息论、语言学和符号学等诸般领域概念。

本泽信息理论美学的代表性观点是他对“自然诗”（“Natürliche Poesie”）与“人工诗”（“Künstliche Poesie”）的区分。^⑤本泽在其代表作《美之编程》

① Theo Lutz, „Stochastische Texte“, S. 4.

② Hermann Rotermund, *Keine Anrufung des großen Bären. Max Bense als Wegbereiter für konkrete Poesie und Netzliteratur*, 2001, https://www.stuttgarter-schule.de/01_08_20bense_rb.html [2019-05-27]

③ Saskia Reither, *Computerpoesie. Studien zur Modifikation poetischer Texte durch den Computer*, Bielefeld: transcript Verlag, 2003, S. 66.

④ 本泽也创作过实验诗歌，代表作如1970年发表的“数字诗”（Zahlgedichte）《第二自我》（“Das zweite Ich”）：“Das zweite Ich” / 1+2+4+71+142 / = / 220/1+2+4+5+10+20+11+22+44+55+110 / = 284 / 1+2+4+71+142。本泽对这首“数字诗”的解释是：“第一个数字220的加数是第二个数字284的分母，而284也是由分母组成，这些分母又是第一个数字的加数。柏拉图将这样的数字称为‘合得来的数字’。而且据说他还将‘合得来的数字’称为‘第二自我’。”

（Max Bense, *Ausgewählte Schriften*, hrsg. von Elisabeth Walther, Band 4: *Poetische Texte*, Stuttgart: J.B. Metzler, 1998, S. 317）这类充满“数字魔法”的诗歌，被实验诗人认为肇端于中国的伏羲（《易经》），该传统一直流传到歌德，比如《浮士德》第一部里女巫所唱的“九九口诀歌”（参见歌德《歌德文集》[第一卷]，绿原译，人民文学出版社，1999年，第75-76页），读者阅读这类诗歌的过程基本就是一个猜谜解码的过程。

⑤ 本泽的这种对于“自然诗”与“人工诗”的划分与诺瓦利斯浪漫主义诗学传统中对“自然诗”（Natürliche Poesie）与“人工诗”（Künstliche Poesie）的定义不尽相同。二者诗学对比参见Reinhard Döhl, „Experiment und Sprache“, in *Ist zwüvel herzen nächgebür. Günther Schweikle zum 60. Geburtstag*, hrsg. Rüdiger Krüger, Jürgen Kühnel und Joachim Kuolt, Stuttgart: Fay, 1989, SS. 367-368.

(*Programmierung des Schönen*, 1960)、《文本理论》(*Theorie der Texte*, 1962)以及《信息理论美学概论》(*Einführung in die informationstheoretische Ästhetik*, 1969)等著作中都论及“自然诗”与“人工诗”的区别,当时在学界引起过广泛讨论。本泽认为,诗歌(Poesie)是对主体性、声音形式以及意义构成的表达,“自然诗”与“人工诗”之间的主要区别在于文本的生成方式上,尽管文学作品的美学价值源于其创作过程,但文学文本的生成方式对于文本审美功能的发挥也具有重要意义。^①本泽认为,“自然诗”以黑格尔所谓“个人诗意”为前提,是一种具有经验、感觉、记忆、思想和想象力等观念意识的诗歌,是对先验世界的语言表达。只有在这种本体论框架中,才能有抒情的自我或虚构的诗歌世界。从这种意义上说,诗意是寓于符号中的信息传递,而这些符号的总体就是诗歌语言,这种语言具有自我关系(Ichrelation)与世界角度(Weltaspekt)。在自然诗中,写作本身就是本体的延续,通过这种延续,文字的世界角度与自我相关联,诗中表达的每一个词都遵循自我的世界经验,其赋予主体的审美地位也可看作是对世界的一种反映。^②所谓“人工诗”则是机器创作的诗歌,不论在当时还是当下都属于前卫实验文学,是一种程序化的语言形态,或可被称为“技术诗”。^③人工诗没有“个人诗意”,也没有先验世界,写作也就不再是本体的延续。因此,在人工诗中既没有一个抒情的自我,也不存在一个虚构的世界(*Theorie*: 143)。人工诗来自计算机作诗程序通过选择性原则将某些参数应用于事先“喂”好的有限的语料库,因此,计算机程序和语料构成了人工诗的基础,其诗歌的生成“与美学相吻合,而并不考虑语义内涵”,而且由于缺少自然诗那源自人类诗人的“有意的开端”,人工诗中的“所有语汇都是天然平等的”(Theorie: 145)。

如果说自然诗的特点在于诗歌创作过程始于一个“有意的开端”,那么人工诗则只有一个“语料的源头”(Theorie: 143)。因此,自然诗可以被阐释且只有通过阐释才能理解其传达的语义信息;而人工诗则只能通过美学信息发挥审美功能。为此,本泽发展出一种诗歌研究方法,对文本中的词汇进行统计评估,并根据列叙来确定其美学价值。本泽使用这种方法研究两种文本形式(人工诗和自然诗)的审美功能,评估其美学质量,借此证明文本形式在美学质量方面并不起决定性作用;也就是说,在创作过程中,人类诗人主观意识的作用并不高于机器创作的无主观性,因此,自然诗并

① Christiane Heibach, *Literatur im elektronischen Raum*, Frankfurt a. M.: Suhrkamp Verlag, 2003, SS. 111-112.

② Max Bense, *Theorie der Texte. Eine Einführung in neuere Auffassungen und Methoden*, Köln: Kiepenheuer & Witsch, 1962, SS. 143-144. 后文出自同一著作的引文,将随文标出该著名称首词和引文出处页码,不再另注。

③ Max Bense, *Einführung in die informationstheoretische Ästhetik. Grundlegung und Anwendung in der Texttheorie*, Hamburg: Rowohlt, 1969, S. 109.

不比人工诗更具美学价值 (*Theorie* : 143)。相反, 由机器生成的人工诗是一种不矫揉造作的、“沉湎于偶然性”的语言; 而自然诗则由于其总是从个人主观意识出发因而是一种被过滤过的、带语义倾向性的语言。由于人工诗并非由一个“有意的开端”而产生, 并不包含“通常意义上的语义载体”, 因此, 人工诗不具有语义信息, 只具有审美信息, 而“与语义信息相反, 审美信息不能被编码, 只能被实现” (*Theorie* : 146)。显然, 本泽将“自然”一词分配给了具有主观性的人, 将“人工”一词分配给了无主观性的机器。自然诗的作者是以语法、语义编码的方式遵循“个人诗意”排列词汇的诗人, 而缺乏作者主观意图的人工诗这种语言形式, 则可以将诗歌从那操控一切的“个人诗意”中解放出来。借助机器, 诗歌与作者分离, 成为一种独立于人的主观性之外的存在。可以说, 人工诗的生成方式更接近一种取代人类主观意图的诗歌创作过程, 人工诗因此有可能成为一种纯粹而绝对的诗歌。

德国斯图加特诗派对混合形式和模糊边界的追求从某种意义上讲也促进了艺术表现领域的扩展。文学与技术之间的这种结合在接下来的几十年中不断加强, 作为机器诗歌实验的先锋, 斯图加特诗派为后来的电子诗歌和机器诗歌的发展做出了决定性贡献。人类机器文学的这个起点在当今人工智能文学蓬勃发展的时代几乎已经被完全遗忘。尽管有人认为计算机在作诗能力上可能永远不会超过人类诗人, 也有人将机器诗歌排除在文艺理论研讨范围之外, 但是, 我们可以肯定地说, 卢茨、本泽和斯图加特诗派不仅在机器诗歌创作方面进行了人类文学史上的首次尝试, 更开创了机器诗歌美学创建的先河, 他们的文学实验和美学创建充分证明了机器诗歌绝不仅仅是一种“词语的游戏”。本泽的信息理论美学为文学艺术与信息技术的结合奠定了美学基础, 它所确定的机器文学的美学特征至今仍然可以在最新的机器诗歌作品中得到印证。

作者单位: 北京大学外国语学院德语系
责任编辑: 龚璇