

外国文学大数据与自主知识体系断想

陈众议

内容摘要：第四次工业革命浪潮滚滚而来，建立在大数据和大模型基础上的人工智能迭代发展，一日千里。然而，反观语言文学学科便不难发现，我们几乎没有像样的数据库。外国语言文学学科更是如此，从经典作家作品，到有关研究资料，皆散沙般存在于各种平台，且大多数著作尚流于纸质载体。与此同时，语言文学无用论甚嚣尘上。于是，与自主知识体系密切关联的两个问题便难以回避：一、轻视语言文学如何传承和发展文化？二、缺乏大数据如何在 AI 时代创建自主知识体系？

关键词：外国语言文学；大数据；自主知识体系

作者简介：陈众议，中国社会科学院学部委员、湖南师范大学特聘教授，主要从事西语文学和比较文学研究。

Title: The Big Data of Foreign Literature and A Chinese Intellectual System

Abstract: As the Fourth Industrial Revolution is rolling in, the iterative development of AI based on big data and big models is advancing rapidly. However, when we look back at the field of language and literature, it is not difficult to find that we hardly possess any comprehensive databases. This issue is particularly pronounced in the field of foreign languages and literatures, where classic works by authors and relevant research materials are scattered across various platforms, with most still confined to print media. Simultaneously, the notion that language and literature are useless is rampant. Consequently, two questions, closely related to the establishment of a Chinese intellectual system, become unavoidable: first, how can we inherit and develop culture by devaluing language and literature? Secondly, how can we establish a Chinese intellectual system in the AI era without big data?

Key words: foreign language and literature; big data; Chinese intellectual system

Author: Chen Zhongyi is a member of the Chinese Academy of Social Sciences (Beijing 100732, China) and a Distinguished Professor at Hu'nan Normal University (Changsha 410081, China), specializing in Hispanic literature and comparative literature. Email: chen-zhy@cass.org.cn

DOI:10.19915/j.cnki.fl.s.2024.0064

先说文化传承与发展问题。它们既是文明互鉴的产物，也关乎民族文化自身的连续性和生命力。从古代“东学西渐”到近现代“西学东渐”，从佛教的传入到“四大

发明”的传出，等等，文明互鉴所在皆是，而外国文学学科无疑是其最新最好的润滑剂。至于民族文化自身的连续性和生命力，我们的四书五经和历代经史子集已经说得很多。譬如《周易》中的阴阳虚实辨正和“观乎天文，以察时变；观乎人文，以化成天下”，^①最终实现天地人和。其在我们满族同胞的完美概括，便是故宫天人之和“太和（殿）”、人人之和的“中和（殿）”及人己之和的“保和（殿）”。这里需要强调的一点是使中华文化得以生生不息的中文。它是世代中华儿女的共同创造，它承载了所有侠义和广义的中华文化并使之发扬光大。钱锺书和杨绛伉俪说过，他们之所以回国，且没有去台湾的唯一理由是中文，因为中文的根在大陆。^②从浩如烟海的图书，到不断发掘的简帛，绝对令人叹为观止！当然，其中还有中外语言文学并驾齐驱的归去来兮。

首先，中文是全中国各民族共同的财富、共同的创造，是古来盛世像统一度量衡一样有意化合、协和的产物。其次，语言作为思维、想象、记忆和交际的原动力和原载体，是一切物质和精神生产的第一要素。因此，除了故友柏杨在《中国人史纲》中关于乡土如何筑牢中华文化、并将我们“束缚”，我必须强调中文这个更大更强的稳定剂和粘合剂。没有后者，也许我们早就被化了。语言的被化恰恰是其他几个古老文明断裂的主要症结所在。试想，如果不是因为中文的海纳百川，及其固有的、无与伦比的辨识度、会意性、象征力和无所不能的超级魔方效应，那么即使它没有被辽化、金化、蒙化、满化，也难免被拉丁化。后者正是一些围绕在《新青年》周围的激进派所鼓吹的。大先生鲁迅在《阿 Q 正传》序言中有过这样的检讨：“写他为阿 Quei，略作阿 Q。这近于盲从《新青年》，自己也很抱歉。”^③

中文从来包括语言和文学，两者结伴而生、须臾不可分离。用最简明的话说，语言是体，文学是魂，二者保证民族心性与审美价值的延续性和生命力。说到延续性和生命力，我们难免想起人文学界挂在嘴边的箴言：“向死而生”（Daseinzumtode）。它是海德格尔提出的一个著名的存在主义命题，是德语“das Sein zum Tode”的集合，出现在其代表作《存在与时间》中，用以讨论人类以及一切存在的终极问题。^④虽然海德格尔没有明确道出生命意义安在，但他的命题策应了文学存在的一个重要理由：理想主义。生命本无常，而且在时间的长河中注定短暂、虚妄，但理想的光芒赋予它以价值和意义，譬如古来救人民于水火的英雄豪杰、国族脊梁，成就利在千秋功在当代的伟大事业者，以及留下荡气回肠感天动地诗章墨宝的文人骚客，等等，等等。

先说文学这个吃不得穿不得的“玩意”，现代以降最令人动容，甚而振聋发聩的是胡适对陈独秀的评价，即后者对五四“文学革命”做出的三大贡献之由我们的玩意儿变成了文学革命，变成了三大主义。^⑤同样，当我们回想起历代爱国诗人的铿锵诗行或侠义文学，家国情怀和为民请命、为国捐躯、见义勇为、大义凛然的豪情壮志便会充盈心扉，乃至令人血脉偾张、热血沸腾。

当然，文学也不全是真善美，古来中外文苑中假恶丑充斥者并不罕见。而这正是文学这个偏正结构中文字赖以存在的最大理由。文学批评不仅需要甄别真善美和假恶丑，而且需要在一般好坏对错之间规避排中律和非好即坏、非白即黑的形而上学，并对其中广袤的混沌区给出合理的和恰当的厘清、甄别与评判。这也是关键时刻文学这

只双翼鸟、这辆两轮车峰回路转、踽踽前行的基本保证。至于我等形形色色两足动物的理智与情感，也是一对须臾不可分离的矛盾统一体。前者是精神之内核，后者是精神之华彩。前者赋予人类以意识和慧心、意志和理性，后者让人类的意识和慧心、意志和理性在审美和感官维度上得以彰显。因此，人类的世界观、人生观和价值观，以及道德感和伦理思想无不浸淫其中。

理智告诉人类向死而生，同时每每有七情六欲妄图将其改变。人工智能和基因工程的发展让无数人膨胀了长生不死的欲望，但反过来这种欲望可能一夜之间毁灭人类。试想，人生几十年尚且如此贪婪，一旦获得长生不死的密钥，那么等待人类的无疑是有违天理、万劫不复。因此，生与死、理智与情感的辩证关系，正是中外文学经典孜孜不竭的自我追问，也是文学理想主义基因的不懈追求。所谓一阴一阳为之原道，一生一死为之生灵。^⑥《易经》讲变化与平衡，犹如《道德经》谓“持而盈之，不如其已”。尊重自然规律，乃是天人合一的真谛。

依此类推，不偏不倚、中正和谐也是理智和情感平衡的不二法门，况如今早已不是简·奥斯汀的时代，科技理性正在资本的助推下所向披靡，人类需要堂吉诃德+哈姆雷特，即既要该出手时就出手，又要对现存和面前的世界适当怀疑态度。这就更需要想象力，需要理智和情感的高度协同。

二

再说大数据，它是数据+价值和审美加工的产物。撇开语言文学，后者便无从谈起。此外，如果我们再不重视大数据建设，那么我们将很快对美西方出现数字赤字、数字鸿沟。也许有人会问：为什么我们不能直接挪用美西方的数据库？答案也许很简单：其一是美国已经对我们关闭 ChatGPT；其二是大数据必须经过价值加工，有价值，甚至审美取向，譬如 SORA 所自的“东京街头某女郎”或“世界末日幸存者”不是极度西化的美女，便是金发碧眼的洋人。

这就是说，我们亟需建设具有主体性和中国心的大数据。后者需要人心根基和家国情怀。她们根植于我国传统文化。一如数理化不分家，文史哲不分家同样是中华文化的优秀传统。文学作为“特殊的意识形态”，外国文学不仅是消遣和情感交流的工具，它也是各民族的审美对象和集体无意识表征；历史也不仅是记忆，而且同是社会的发展过程和物质生产、社会交往，甚至和文学一样是各国族集体意识、家国情怀的主要来源。哲学就不必说了，它业已进入“101计划”优选学科。但是，需要强调的是，当我们在诵读唐诗宋词时，鲜有不在内心产生家国情愫的。这是由于优秀的语言文学与民族心理之间难分难解的亲缘关系。盖因文学归根结底是民族理想与情感、价值与审美等多重精神要素的集结。而中文永远是最大的传统。问题是，我们在做些什么？从幼儿到研究生，语文教育退化严重，以至于文科博士生不擅写作，遑论文采飞扬。这当然不是外语之过。想当初清华四大导师，其中文外文是相得益彰的。

再说文学的秘诀，或可称之为文心者，往往首推情感。有道是情感乃人类最高级生命体验，犹如太阳，让审美、意义、价值、伦理、修辞等一众行星远近环绕。一切

生灵的是非好恶、喜怒哀乐皆在经意和不经意中服从于她。一个怦然心动或毛发森竖，抑或喜极而泣、潸然泪下，辗转反侧、夜不能寐，常常取决于“有感而发”。优秀的文学作品大抵都是有感而发，那么文学批评呢？是否也应该有感而发，带着脉动、体温和风骨？！

怪谲的是现代西方文论鲜有关注读者情感诉求的，即或有也已然被林林总总“高大上”的主义和后主义所掩埋。一个执着于文学“教育”的文本细读或一些致力于叙事形式的砖砖瓦瓦抹去了情感作为审美核心的亘古逻辑。当然，我不是说文本细读和叙事形式不重要，但它们必须服从文心的需要。诚所谓文无心不美、无情不丽，刘勰在《文心雕龙·情采》中写道：“夫桃李不言而成蹊，有实存也；男子树兰而不芳，无其情也。夫以草木之微，依情待实，况乎文章，述志为本，言与志反，文岂足征？”^⑥钟嵘在《诗品·序》中谓“气之动物，物之感人，故摇荡性情，形诸舞咏。照烛三才，晖丽万有。灵祇待之以致飨，幽微藉之以昭告天地，感鬼神，莫近于诗。”^⑦

惊天地、泣鬼神，但前提是诗文有魂有情。非如此，文心不能如阳光般晖丽万物，烛照天、地、人类。

作为结语，我想说上千年、上百年，尤其是近四十多年的中外交流早已使我们融入了世界大家庭和人类命运共同体。中文也早就不是钱玄通等五四新青年所曾经理解的“固化”的方块字。文字之妙，妙在承载文化，进而铸造魂魄、改善世道人心。同时，文明因互鉴而扬长补短、丰饶充盈。从“百日维新”到五四运动再到改革开放，几代外语人才不仅引进了德先生、赛先生，还有马克思主义。由于我党将马克思主义同中国实际和悠久文化相结合奠定了基础，并反哺和丰富了中华文化，使之焕发出令全世界刮目的灼灼光辉。如今，新时代中国特色社会主义建设必须依仗文史哲等为核心的人文社科，并且博采中外，否则大数据、大模型无从谈起。归根结底，人工智能主要是算法加数据，而后者恰恰是建立在语言文字基础上的文史哲和数理化等全部人文社科和理工类大数据。因此，作为文史哲等侠义文化基础的中外语言文学如果得不到重视、不能作为优先发展学科，那么同心圆式有心有园的大模型将付之阙如。

注解【Notes】

- ①⑥《周易》，中华书局，2010年，第94、290页。
 ②杨绛：《我们仨》，三联书店，2003年，第122页。
 ③鲁迅：《鲁迅全集》第1卷，人民文学出版社，2005年，第514页。
 ④参见海德格尔：《存在与时间》，陈嘉映、王庆节译，三联书店，2014年，第275—276页。
 ⑤胡适：《胡适学术文集·新文学运动》，中华书局，1998年，第192页。
 ⑦《道德经》，中华书局，2006年，第22页。
 ⑧刘勰：《文心雕龙·情采》，王运熙等译注，上海古籍出版社，1998年，第287页。
 ⑨钟嵘：《诗品·序》，人民文学出版社，2009年，第1页。

责任编辑：王树福