

殷墟卜辞族名、地名与方国名的计算机辅助判别新探： 基于大语言模型的多智能体规则推断框架*

杨清越¹ 孙亚冰² 赵清荣³ 武笑迎¹ 刘 艺⁴ 张 愉⁴

1. 中国社会科学院考古研究所, 北京 100101 2. 中国社会科学院古代史研究所, 北京 100101

3. 殷墟博物馆, 河南 安阳 455000 4. 澳大利亚格里菲斯大学 (Griffith University)

Abstract: Traditional research on the historical geography of Oracle Bone Inscriptions (OBI) relies heavily on manual interpretation following the “Two-fold Evidence” approach. While effective, this method is constrained by limited efficiency and inevitable subjectivity when processing large corpora, particularly in resolving challenges such as the integration of clan and place identities and the ambiguity created by overlapping names. This study presents an automated framework for identifying proper names in OBI based on Large Language Models (LLMs) and a Multi-Agent System (MAS). Expert knowledge, including syntactic patterns, generic name markers, and contextual evidence, is formalized into structured prompts that guide five specialized agents: Presence Detection, Place Name Extraction, Clan Name Extraction, Polity Name Extraction, and Consistency Verification. The framework further incorporates the Reflexion mechanism to enable iterative logical self-correction. Experiments conducted on approximately 16,000 oracle bone inscriptions show that the system achieves 92.3% accuracy in presence detection, with identification accuracies of 87%, 83%, and 79% for place names, clan names, and polity names, respectively, substantially outperforming conventional rule-based approaches. The results demonstrate that LLMs possess strong semantic disambiguation capabilities even in the low-resource context of ancient scripts. Beyond enabling efficient large-scale screening of OBI corpora, the proposed framework offers a promising computational approach to reconstructing the political geography of the Shang Dynasty and supporting the development of historical knowledge graphs.

Keywords: Oracle Bone Inscriptions (OBI), Named Entity Recognition (NER), Large Language Models (LLMs), Multi-Agent Systems (MAS), Shang Dynasty historical geography

摘要: 甲骨文地理研究传统上依赖“二重证据法”进行人工考释,但在处理海量语料时面临效率瓶颈与主观性偏差的挑战,尤其是“族地合一”与“名号重叠”的辨别仍是学术难题。本文提出一种基于大语言模型(LLM)与多智能体系统(MAS)的甲骨卜辞专名自动识别框架。该框架将甲骨学专家知识(如句法分布、通名标志、活动语境)形式化为结构化提示指令,构建了由存在性判定、地名识别、族名识别、方国名识别及一致性核验组成的五个独立智能体,并引入Reflexion机制实现逻辑自纠正。对约1.6万条卜辞的实验结果显示:系统在存在性判定上的准确率达92.3%,地名、族名、方国名识别准确率分别为87%、83%和79%,显著优于传统正则匹配方法。研究表明,大语言模型在低资源古文字语境下具有强大的语义消歧能力,不仅能有效辅助大规模语料的自动化初筛,也为重构商代政治地理格局及构建历史知识图谱提供了智能化路径。

关键词: 甲骨文;命名实体识别;大语言模型;多智能体系统;商代地理

*本文为中国社会科学院考古研究所创新工程项目“人工智能在考古学中的应用”(项目编号:2026KGYJ007)和腾讯探元计划NextGen项目“契文定域·智绘山河——基于多智能体和甲骨释文的商代地理空间的重现(项目编号:R0100051)”阶段性研究成果。

自孙诒让^[1]、王国维^[2]开启以卜辞线索勾勒商代疆域轮廓的先河以来,甲骨地理研究便在“二重证据法”的指引下不断向精密化演进。郭沫若^[3]、董作宾^[4]、陈梦家^[5]等前辈学者通过严谨的文字考释与文献互证,逐步构建起商代地理的宏观图景。然而,在微观判别层面,一个长期存在的学术难题是“族地合一”与“名号重叠”的辨别。

以《甲骨文合集》(本文简称《合集》)第20500条“戊戌卜:雀人芻于效”为例,辞中“雀”字究竟指代血缘组织(族名)还是地理空间(地名)^[6],学界至今仍存分歧。分歧并非偶然,它反映了商代社会“因人名地、族地同名”的本质特征:族群依血缘聚居成邑,使得同一字形在不同语境下具备了“族、地、人”多重属性的动态重叠^[7]。

学界针对此类问题的判别,主要形成了三类技术路径。文献比勘法:依托传世文献如《尚书》《左传》等进行名实比对^[8],其局限在于商周之际地名沿革剧烈,文献记载往往存在“时代错置”与“视角偏差”;干支系联法:利用干支纪日的连续性,系联同一贞人或王室成员的活动路径^[9],此法对复原田猎路线极具价值,但受限于卜辞残缺导致的逻辑链中断,难以覆盖缺乏干支系联线索的孤立地名;系统内证法:通过归纳介词(如“于”“自”)后的句法分布及通名标志(如“邑”“田”)进行语义锁定^[10],这代表了甲骨学最严谨的内证逻辑,亦是目前判别的主流准则。

尽管上述方法已臻成熟,但在处理大规模语料时却面临着挑战。首先,认知负荷与处理效能的不对称:面对十多万片甲骨带来的海量文本,依赖“专家直觉”的人工判别本质上是一种低效的单线程劳动,难以实现多重规则在万量级数据上的判别和复核;其次,判别尺度的非均质性:由于缺乏形式化的逻辑约束,不同研究者在处理复杂语境时,容易陷入主观心证的循环论证,可能影响研究结论的稳定性。这些情况促使我们思考,能否将甲骨学积累的深厚判别知识进行形式

化迁移,利用计算机辅助程序在保持学术严谨性的同时,实现大规模语料的自动化属性初筛与消歧。

一 计算机辅助判别的可能性

(一) 自然语言处理与命名实体识别

命名实体识别(NER)旨在从非结构化文本中识别并分类特定类型的实体,如地名、人名等^[11]。尽管NER技术在现代语言中已经比较成熟,但应用于甲骨卜辞时仍有挑战。甲骨文缺乏词边界标记,标注语料匮乏,且存在大量异体字或未释字,使得依赖大规模数据的传统监督学习模型(如BERT或BiLSTM-CRF)无法直接迁移^[12]。

大语言模型(LLM)的兴起为解决此类低资源、高复杂度的NER任务提供了新路径。以GPT-4为代表的模型依托Transformer架构的自注意力机制,能够动态计算文本中各词元间的关联权重,从而精准捕捉长距离的语法依赖与语义逻辑^[13]。

不同于传统微调范式,大语言模型可通过“提示工程”(Prompt Engineering)在零样本或少样本条件下执行任务^[14]。本研究的核心在于将甲骨地理学界积累的句法规则、通名标志等领域专家知识,直接转化为自然语言提示指令。这种方法有效地将学界共识外化为可计算的判定规则,使模型能在大规模语料中执行学术化的语义推断,充当高效的初筛与判别工具^[15]。

(二) 判别规则的形式化策略

将传统甲骨学的考释经验转化为大语言模型可解析的形式化指令,是实现学术知识计算化的核心环节。这一过程旨在解决知识的“逻辑确定性”,即将学界公认的判定条件转化为可执行的算法规则。

本研究在参考郭沫若、陈梦家、郑杰祥、孙亚冰^[16]等学者的实证准则基础上,采用了“结构化提示工程”(Structured Prompting)策略^[17]。其核心逻辑在于将复杂的语义辨析拆解为三个层

面的“算子”：句法分布算子，锚定介词短语与动宾结构的位标特征；范畴特征算子，集成建筑物、地形及聚落等通名作为指征；语义推断算子，耦合田猎、军事、祭祀等宏观语境的逻辑指向。

为确保识别系统的可靠性，本研究引入了基于Reflexion框架的一致性核验机制^[18]。通过约束模型输出标准的JSON格式，并设置逻辑回溯链条，系统能够对识别过程中可能出现的属性冲突进行自发性重新判别。这种设计使模型不再仅仅是词频统计工具，而成为了一个遵循甲骨学逻辑的“机器考释员”。

二 判别框架设计与技术路径

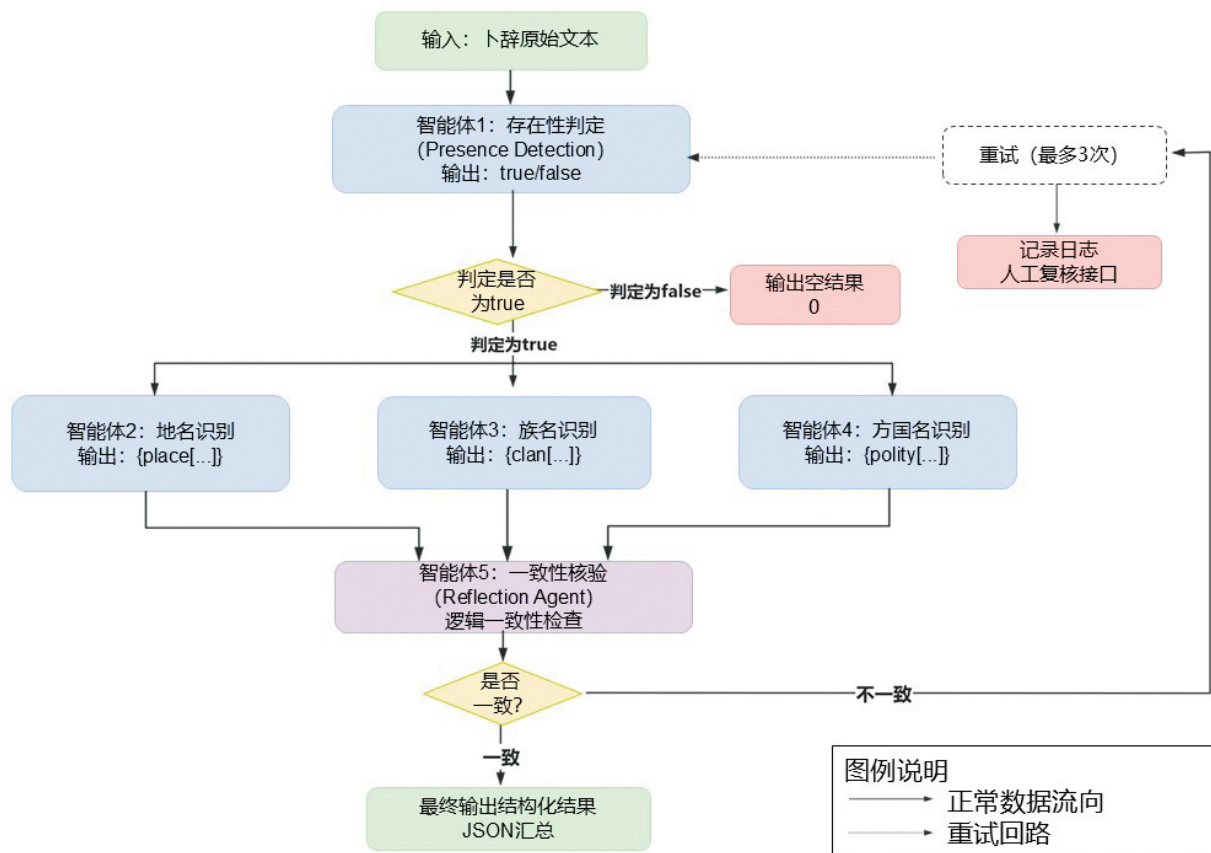
本研究采用多智能体系统（Multi-Agent System, MAS）架构^[19]，将复杂的专名判别任务解构为五个功能独立的子任务。（图一）每个智能体封装了特定的领域知识与判别算子，通过预

设的编排协议（Orchestration Protocol）进行动态调度。系统底层基于AsyncIO异步框架驱动并发处理^[20]，并调用GPT-4o模型执行推理。相较于单智能体方案，MAS架构能有效规避提示词过长导致的“注意力稀释”问题，并支持无依赖提取任务的并行执行，显著提升了处理效率。同时，系统引入独立的核验智能体构建闭环反馈机制，确保了不同智能体输出结果的一致性^[21]。以下分述五个智能体的技术设计与实现细节。

（一）智能体1：存在性判定

存在性判定智能体（Presence Detection Agent）执行二分类任务，判定输入卜辞中是否包含地名、族名或方国名中的至少一种。该智能体接收原始卜辞文本作为输入，输出约束为布尔值“true”或“false”。

这一前置过滤机制实现了计算资源的条件分配。通过三类判别信号：句法标志、通名标志以及活动语境标志，剔除不含专名的卜辞，将大规模语料处理的API调用成本显著降低，提升了工



图一 地名、族名和方国名识别多智能体系统

程效率。

(二) 智能体2：地名识别

地名识别智能体 (Place Name Extraction Agent) 在存在性判定为true的条件下被激活，负责识别并提取地理空间实体。核心判别规则有以下两条。

第一，句法分布规则。地名紧接介词“于”“自”“从”“在”“至”等之后，或作为动补短语置于动词之后，或跟随动词的地点宾语^[22]。

第二，通名辅助规则。当专名后伴有建筑物类（宅、庭、门、京、宫、束、官等）、自然地理类（田、兆、山、麓、泉、丘等）、城邑区域类（邑、亭、囿、墉等）^[23]通名成分时，地名属性的置信度显著提升。

智能体2特别关注与商王核心活动相关的四类地名^[24]：田猎活动所涉地望，军事征伐中涉及的敌对方国名，与宗庙祭祀密切相关的礼制场所，以及与农业生产相关的耕作区域。

识别结果输出为JSON格式：{“place”：[“地名1”，“地名2”，…]}。若未检测到地名，则返回空JSON对象{}。这种结构化的输出保证了系统内部数据传输的准确性，也方便后续进行大规模的自动化处理与统计分析。

(三) 智能体3：族名识别

族名识别智能体 (Clan Name Extraction Agent) 专门负责识别族群身份及血缘组织名称，与智能体2并行运行，其提示指令中编码了多条件的逻辑表达式^[25]。

基础判别条件：1. 称谓规则，识别“某子（子某）”“某族”“某侯”“某人”等族群命名形式；2. 活动规则，分析语境是否以族称为行为主体或客体，涉及征伐、田猎、祭祀等活动；3. 区域规则，考察该名号在卜辞中是否常与特定地理空间关联，或其本身即兼作地名。

辅助佐证条件：1. 贡纳规则，考察是否存在关联物产并见纳贡记录；2. 军事记录，是否为征伐卜辞中被明确提及的主体或客体；3. 铭文支持，相关人名是否出现在青铜器族徽铭文中。

通过基础条件与佐证条件交叉印证，提升判

定的置信度。针对“族地合一”引发的语义重叠，智能体3重点结合上下文纳贡行为、军事动员等动态语境特征，判断该字在当前语境中是指代血缘集团还是空间范畴。

智能体3输出格式为 {“clan”：[“族名1”，“族名2”，…]}。

(四) 智能体4：方国名识别

方国名识别智能体 (Polity Name Extraction Agent) 负责识别商王朝周边的方国或区域名称。与智能体2和智能体3并行，其提示指令聚焦于两个判别维度。

第一，直接称谓标志。当卜辞中出现“某方缺”形式时，可直接判定为方国名。

第二，间接政治语境关联。当卜辞中出现“某侯”且上下文涉及征伐、朝贡等方国互动语境时，该名称亦被纳入方国名候选。智能体需依赖地缘政治语境的深层分析，尤其是考察其与商王朝的互动关系性质。

智能体4输出格式为 {“polity”：[“方国名1”，“方国名2”，…]}。

(五) 智能体5：一致性反思与结果核验

一致性反思智能体 (Reflection Agent) 是整个系统的质量保障环节，借鉴了Shinn等人提出的Reflexion框架^[26]。智能体5接收所有智能体的输出结果作为输入，执行以下核验逻辑：若智能体1的存在性判定为true，则智能体2、3、4中至少一个应返回非空列表；若智能体1的判定为false，则智能体2、3、4的输出均应为空列表。

智能体5的输出是“一致”或“不一致”。当核验结果为不一致时，系统触发自纠正机制，重新执行判别流程，最多允许三次重试迭代。若三次重试后仍不一致，系统保留最后一次的输出结果并记录日志，留待人工复核。

针对“未释字”现象，系统采取保真策略，将其视为独立语义单元原样返回，为后续专家复核预留接口。

(六) 系统调度与工程优化

在系统调度上，本研究采用“串行过滤—并行提取—串行核验”的混合模式。为平衡大规模

中原文物
CULTURAL RELICS OF CENTRAL CHINA

语料处理的效率与成本，系统基于 AsyncIO 异步框架，以 10 条卜辞为单位进行批处理滚动。针对网络波动与模型推理的随机性，系统内置了双层容错方案：底层执行网络层面的自动重试（上限 3 次），上层配合 Reflexion 框架执行语义层面的逻辑自纠正（上限 3 次）。在处理一万六千条卜辞（API 总调用约 6 至 10 万次）的过程中，该架构有效平衡了工程稳定性与计算成本^[27]。

三 实验评估与判别规则的逻辑验证

（一）评价指标体系和数据来源

本研究采用命名实体识别（NER）通用的准确率（Precision）、召回率（Recall）及 F1 值作为核心评价指标^[28]。针对前置的存在性判定任务，则辅以分类准确率（Accuracy）进行衡量。鉴于甲骨文专名存在异体、省写等形态变体，在实体匹配逻辑上，本研究采用了双向子串匹配（Bidirectional Substring Matching）策略。即当模型提取结果与参考条目互为子串关系时，均判定为匹配成功^[29]。这一策略旨在更客观地反映系统在处理非规范文字形态时的实际识别效能，有效避免了因精确匹配导致的效果低估。

本研究使用的数据来源于甲骨文数字化工程的释文数据库，共包含约一万六千条经过预处理的甲骨卜辞文本。这些卜辞主要来自《合集》《甲骨文合集补编》《小屯南地甲骨》等著录。数据预处理包括去除重复条目、清洗格式异常的文本、统一字符编码等。原始数据总量为 19,580 条，去重后得到有效文本约 16,000 条。参考数据集包括三个权威列表：地名列表 89 条，族名列表 23 条，方国名列表 21 条，主要参考郑杰祥《商代地理概论》、钟柏生《殷商卜辞地理论丛》、孙亚冰及林欢《商代地理与方国》进行编制^[30]。这些列表仅包含学界已有定论的专名，对于尚存争议的专名暂不纳入。识别效果的评价以人工标注的测试集为基准，由两位专家独立标注后取一致意见。

（二）识别效果

验证结果表明，本系统在甲骨卜辞专名判别

方面取得了较好效果，各智能体正确率见表 1。

存在性判断的准确率达 92.3%，表明系统能有效筛选出包含专名的卜辞。地名识别正确率最高（87%），这与地名的句法标志较为显著有关——介词“于”“自”“在”与地名的搭配模式辨识度较高。族名识别次之（83%），其判别所依赖的称谓模式和活动语境特征较为多样。方国名识别正确率相对较低（79%），原因有二：一是方国名在卜辞中出现频率远低于地名和族名，模型可参照的语境样本较少；二是参考列表中方国名仅 21 条，个别误判对整体正确率的影响较大^[31]。

核验一致率 89.6% 表明，绝大多数卜辞在首次处理后即通过了一致性检验，无需触发重试机制。

（三）基准方法与对比实验

为量化评估本文研究方法的增益效果，本研究设计了两组基准方法进行对比。

1. 规则匹配基准（Rule-Based）。利用正则表达式实现本文所述的语言学规则。该方法与本文方法使用完全相同的规则集，区别仅在于执行方式——正则匹配仅执行死板的字面匹配，不具备语境理解能力。

2. 随机基准（Random）。按预设概率从列表中随机抽取实体，作为底线参照^[32]。三种方法的对比结果见表 2。

对比分析可见，首先，规则匹配的指标显著

表 1 各智能体正确率判别步骤

各智能体正确率判别步骤	正确率
智能体 1-存在性判断	92.3%
智能体 2-地名识别	87%
智能体 3-族名识别	83%
智能体 4-方国名识别	79%
智能体 5-核验一致率	89.6%

表 2 三种方法对比

指标	随机基准	规则方法	本文方法
存在性判断正确率	~40%	59.8%	92.3%
地名识别正确率	——	52%	87%
族名识别正确率	——	48%	83%
方国名识别正确率	——	41%	79%

高于随机基准,验证了本研究所拟定的语言学规则在捕获专名特征上的有效性。其次,本文方法在所有指标上均大幅超越规则匹配,正确率提升均在30个百分点以上。尤其在方国名识别上,提升最为突出。

这种性能差异揭示了大语言模型在语义消歧上的核心优势,它能够通过语境自动抑制误判。此外,正则匹配难以处理同一字在不同卜辞语境中的语义微差,而多智能体架构通过深层推理有效地解决了这一逻辑难题。

(四) 案例分析

1. 典型地名识别

“贞:王其田于襄,𠂔于河。吉。”(《合集》30439)

程序准确将“襄”判定为地名,其依据主要源于以下三重证据的逻辑交叉验证:第一,句法位置,“襄”位于介词“于”之后,符合卜辞中地名的典型句法位置;第二,从语义逻辑来看,“田于某”构成的动补结构要求其后必须衔接地理方位词;第三,辞例佐证,“襄”地在卜辞中出现频次较高,有“自襄至于大,亡灾。吉。”(《合集》28188)、“王其田襄,射兕,亡灾,侃王。吉。”(《屯南》1098)等辞例,郑杰祥、孙亚冰、唐英杰等学者均对其地望进行了考证^[33]。

2. 复杂语境中的族名与方国名识别

“令多子族累犬侯捷周,堪王事。”(《合集》6813)

程序将“多子”判断为族名、“犬侯”为方伯/方国名。判定逻辑的回溯如下:第一,“多子族”(多子即多位“子某”)^[34]采用了“某族”的典型命名范式;第二,“犬侯”采用了“专名+爵称”的结构,系征伐执行者之一,“犬侯”与“多子族”同辞受命,其臣属方国首领身份获旁证强化,符合方伯的身份标识。

3. 通名辅助识别

“辛丑卜,殷贞:妇姁呼泰于丘商,[受年]。”(《合集》9530)

程序利用通名判定规则准确识别了其地名属性。在“丘商”这一组合词中,“丘”作为自然

地理地名通名,起到了明确的范畴界定作用;而“商”则作为专名,提供了具体的空间标识;卜辞“于丘商”的介宾结构是地名最稳定的句法分布环境,与通名判定形成互证。董作宾、陈梦家、岛邦男、钟柏生、郑杰祥、李雪山、孙亚冰等学者均对其地望进行过考证^[35]。

4. 一字多义识别

“丁巳卜,令雀即雀。”(《合集》20171)

程序将两个“雀”字分别识别为族名和地名。推理路径如下:(1)地名判定依据:动词“即”指向目的地,其后宾语须为地点名词,判断第二个“雀”为地名;(2)族名判定依据:动词“令”为派遣动词,后需接人或人群为宾语,判定第一个“雀”可能指代族群。两条判定分别来自不同规则集的智能体,输出结果逻辑兼容,一致性检验判定为consistent(地名非空,满足存在性为true的约束)。

这一双重标注恰好反映了“雀”的一字多义。模型的地名判定路径与句法规则一致,族名判定路径则与族群活动模式的判别标准吻合。张惟捷论证“雀”由私名演变为族名、再转变为地名的研究指出^[36],族名往往转化为地名,从族名到地名的演变是商代专名系统的普遍规律之一。模型无法判断这一历史演变过程,但在给定规则驱动下,同时捕捉到了“雀”在同一辞例中的地名和族名两种信号,这一结果与学界认为“雀”兼具多重含义的认识一致。

(五) 误判分析

1. 时间词与地名的识别混淆

“癸卯卜,今日雨。”(《合集》12870)

本条卜辞属于询问当日天气的常规记录。程序错误地将“今”字判定为地名。回溯判定过程发现,模型在规则库中习得“今”字在部分辞例中可用作地名,如《合集》第14380条(“癸贞:王至于今水”)等辞例中确实存在作为地名的用法,错误地将先验知识迁移至当前语境。此误判是卜辞专名词义消歧的困难,即同一字在不同语境中承载不同语义功能,加之部分卜辞简短或存在残缺,上下文窗口不足(insufficient context),

中原文物
CULTURAL RELICS OF CENTRAL CHINA

导致模型难以从内在多义性 (inherent polysemy) 中做出正确选择。

2. 族名与方国名的分类边界模糊

“辛酉卜，殷贞：今春王比望乘伐下弁，受有佑。”（《合集》6482）

程序将“望”字同时标注为族名与方国名，产生了类别标签的冗余。此类现象与其说是误判，不如说是族名与方国名共有的特征信号（“涉及军事活动”）高度重合所引发的类别边界模糊问题。学界对“望乘”等特定人物的身份属性尚存争议^[37]，参考标准本身的不确定性导致模型在处理此类复杂语境时，由于缺乏足够区分度的信号，在决策边界上产生了重叠。

四 余论

本研究用计算机程序设计了一套分步流程，辅助判断甲骨卜辞里的族名、地名和方国名。从测试结果看，存在性判断的正确率达到了92.3%，地名、族名、方国名的识别正确率分别是87%、83%和79%，明显高于传统规则方法。

从材料价值来看，甲骨卜辞中这几类专名的系统判别，是重构商代政治地理格局的基础。本研究处理了一万六千条左右的卜辞数据，识别结果经过系统整理，可以为学界提供一批基础材料。有了这批材料，学者能够在大规模样本的基础上重新审视商代方国的地理分布、族群迁徙等问题。自动识别的结果也能为商代地理的进一步研究提供参考。

从研究方法来看，本研究把大语言模型引入甲骨学研究，把传统研究积累下来的判别规则如句法标志、通名规则等系统化。这个过程也是对前人研究成果的一次梳理和检验。传统甲骨学研究中，学者们积累了丰富的判别经验，但这些经验大多以隐性知识的形式存在，很难被系统地提取和共享。此次研究把这些经验转化成了明确的判别规则，给计算机程序提供可执行指令的同时，也给学界留下了一份较为系统的规则体系。

这份体系对初学者来说可以作为入门参考，研究者之间也可以拿它作为交流讨论的共同基础。此外，本研究开发的判别程序作为一种工具，能辅助学者快速筛选大量卜辞数据，提高基础工作阶段的效率。大语言模型推理过程本身也有可解释性、可验证性，能够跟传统甲骨学的考证逻辑形成有效对话。

要说明的是，计算机辅助判别终究是辅助，代替不了人工考释。甲骨学研究的核心，还是学者凭借专业知识和学术判断所做的释读工作。程序的价值在于让学者能把更多精力放在那些真正需要深入思考的核心问题上。当然，本研究也有不少局限，参考数据采集的规模还不够大，判别依据也无法做到完全透明，遇到学界有争议的专名，还需结合具体材料人工复核。展望未来，这项研究还有许多可拓展空间。随着人工智能技术的发展和数字化工作的推进，可以扩大数据规模，让评估结果更扎实，也可以结合甲骨文字形图像识别技术，实现端到端的处理。更有应用前景的方向是，基于自动识别的族名、地名、方国名，构建商代历史地理知识图谱，为学界提供一个可视化、可检索的知识平台。

[1] 孙诒让. 契文举例 [M]. 上海: 蟬隐庐, 1927.

[2] 王国维. 殷卜辞中所见先公先王考 [C] // 观堂集林 (卷九). 北京: 中华书局, 1959: 409-450; 殷卜辞中所见先公先王考续考 [C] // 观堂集林 (卷九). 北京: 中华书局, 1959: 409-450; 王国维. 鬼方昆夷玁狁考 [C] // 观堂集林 (卷十三). 北京: 中华书局, 1959: 583-606.

[3] 郭沫若. 卜辞通纂 [M]. 北京: 科学出版社, 1983; 郭沫若. 殷契粹编考释 [M]. 日本东京: 求文堂石印本, 1937.

[4] 董作宾. 甲骨文断代研究例 [C] // 历史语言研究所集刊 (外编第一种《庆祝蔡元培先生六十五岁论文集》上册). 北京: 商务印书馆, 1933; 董作宾. 殷历谱 [C] // 董作宾先生全集乙编 (第1册). 台北: 艺文印书馆, 1977.

[5] 陈梦家. 殷虚卜辞综述 [M]. 北京: 中华书局, 1988.

[6] 陈梦家. 殷虚卜辞综述 [M]. 北京: 中华书局, 1988: 297-298; 郑杰祥. 商代地理概论 [M]. 郑州: 中州古

- 籍出版社, 1994: 15; 张惟捷. 殷商武丁时期人物“雀”史迹研究 [J]. 历史语言研究所集刊, 2014 (第八十五本第四分); 王建军, 杜佳浩. 殷卜辞所见的雀族及其相关问题 [J]. 中州学刊, 2020 (4).
- [7] 朱凤瀚. 商周家族形态研究 (增订本) [M]. 天津: 天津古籍出版社, 2004: 27-117; 林沄. 从武丁时代的几种“子卜辞”试论商代的家族形态 [C] // 林沄学术文集. 北京: 中国大百科全书出版社, 1998: 46-59; 郑杰祥. 商代地理概论 [M]. 郑州: 中州古籍出版社, 1994: 28.
- [8] 王国维. 殷卜辞中所见先公先王考 [C] // 观堂集林 (卷九). 北京: 中华书局, 1959: 409-450; 王国维. 殷卜辞中所见先公先王考续考 [C] // 观堂集林 (卷九). 北京: 中华书局, 1959: 409-450; 陈梦家. 殷虚卜辞综述 [M]. 北京: 中华书局, 1988.
- [9] 郭沫若. 卜辞通纂 [M]. 北京: 科学出版社, 1983: 7; 钟柏生. 殷商卜辞地理论丛 [M]. 台北: 艺文印书馆, 1989: 30-38.
- [10] 陈梦家. 殷虚卜辞综述 [M]. 北京: 中华书局, 1988: 85-134; 裘锡圭. 谈谈殷墟甲骨卜辞中的“于” [C] // 裘锡圭学术文集·第1卷甲骨文卷. 上海: 复旦大学出版社, 2012: 527-551; 钟柏生. 殷商卜辞地理论丛 [M]. 台北: 艺文印书馆, 1989: 350-351; 赵庆森. 商周族群迁徙与地名变迁 [M]. 上海: 上海古籍出版社, 2024: 67-74.
- [11] Li J, Sun A, Han J, et al. A Survey on Deep Learning for Named Entity Recognition [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34 (1); Sang E T K, De Meulder F. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition [C] // *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL*. Edmonton: ACL, 2003: 142-147; Wang S, Sun X, Li X, et al. GPT-NER: Named Entity Recognition via Large Language Models [EB/OL]. arXiv: 2304.10428, 2023; Xie T, Li Q, Zhang J, et al. Empirical Study of Zero-Shot NER with ChatGPT [C] // *Proceedings of EMNLP*. Singapore: ACL, 2023: 7935-7956.
- [12] Lafferty J, McCallum A, Pereira F. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data [C] // *Proceedings of the 18th International Conference on Machine Learning (ICML)*. San Francisco: Morgan Kaufmann Publishers, 2001: 282-289; Lample G, Ballesteros M, Subramanian S, et al. Neural Architectures for Named Entity Recognition [C] // *Proceedings of NAACL-HLT*. San Diego, ACL, 2016: 260-270; Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [C] // *Proceedings of NAACL-HLT*. Minneapolis: ACL, 2019: 4171-4186.
- [13] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need [C] // *Advances in Neural Information Processing Systems 30 (NeurIPS)*. Long Beach: Curran Associates, Inc., 2017: 6000-6010; OpenAI. GPT-4 Technical Report [EB/OL]. arXiv: 2303.08774, 2023.
- [14] Wang X, et al. Self-Consistency Improves Chain of Thought Reasoning in Language Models [C] // *Proceedings of ICLR*. Kigali: ICLR, 2023; Xie T, Li Q, Zhang J, et al. Empirical Study of Zero-Shot NER with ChatGPT [C] // *Proceedings of EMNLP*. Singapore: ACL, 2023: 7935-7956.
- [15] Wang X, et al. Self-Consistency Improves Chain of Thought Reasoning in Language Models [C] // *Proceedings of ICLR*. Kigali: ICLR, 2023.
- [16] 孙亚冰, 林欢. 商代地理与方国 [M]. 北京: 中国社会科学出版社, 2010.
- [17] White J, et al. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT [EB/OL]. arXiv: 2302.11382, 2023.
- [18] Shinn N, et al. Reflexion: Language Agents with Verbal Reinforcement Learning [C] // *Advances in Neural Information Processing Systems, Vol. 36*. New Orleans: Curran Associates, Inc., 2023.
- [19] Wu Q, Bansal G, Zhang J, et al. AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation [EB/OL]. arXiv: 2308.08155, 2023. 本研究使用的底层模型为 GPT-4o, 通过 OpenAI AsyncOpenAI 客户端进行异步调用。
- [20] [27] Python Software Foundation. Asyncio — Asynchronous I/O [EB/OL]. Python 3.12 Documentation, 2024.
- [21] Du Y, Li S, Torralba A, et al. Improving Factuality and Reasoning in Language Models through Multiagent Debate [C] // *Proceedings of the 41st International Conference on Machine Learning (ICML)*. Vienna: PMLR, 2024.
- [22] 钟柏生. 殷商卜辞地理论丛 [M]. 台北: 艺文印书馆, 1989: 350-351; 葛亮. 甲骨文田猎动词研究 [C] // 出土文献与古文字研究 (第5辑). 上海: 上海古籍出版社, 2013: 31-153; 赵庆森. 商周族群迁徙与地名变迁 [M]. 上海: 上海古籍出版社, 2024: 69-70.

- [23] 宋镇豪. 甲骨文中所见的殷商建筑称名 [C] //甲骨文与殷商史 (新三辑). 上海: 上海古籍出版社, 2013: 5-34; 马盼盼. 殷墟甲骨文所见地名的整理与研究 [D]. 吉林大学博士学位论文, 2022: 26-30; 赵庆森. 商周族群迁徙与地名变迁 [M]. 上海: 上海古籍出版社, 2024: 70-87.
- [24] 陈梦家. 殷虚卜辞综述 [M]. 北京: 中华书局, 1988; 李学勤. 殷代地理简论 [M]. 北京: 科学出版社, 1959; 王玉哲. 中华远古史 [M]. 上海: 上海人民出版社, 2000; 孙亚冰, 林欢. 商代地理与方国 [M]. 北京: 中国社会科学出版社, 2010.
- [25] Xie T, Li Q, Zhang J, et al. Empirical Study of Zero-Shot NER with ChatGPT [C] //Proceedings of EMNLP. Singapore: ACL, 2023: 7935-7956.
- [26] Shinn N, et al. Reflexion: Language Agents with Verbal Reinforcement Learning [C] //Advances in Neural Information Processing Systems, Vol. 36. New Orleans: Curran Associates, Inc., 2023.
- [28] Sang E T K, De Meulder F. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition [C] //Proceedings of the Seventh Conference on Natural Language Learning (CoNLL). Edmonton: ACL, 2003: 142-147.
- [29] 双向子串匹配在处理形态变体丰富的低资源语言NER任务中较为常见, 参见 Tsai R T H, et al. NERBio: Using Selected Word Conjunctions, Term Normalization, and Global Patterns to Improve Biomedical Named Entity Recognition [J]. BMC Bioinformatics, 2006 (S11).
- [30] [33] 郑杰祥. 商代地理概论 [M]. 郑州: 中州古籍出版社, 1994: 199-202; 孙亚冰, 林欢. 商代地理与方国 [M]. 北京: 中国社会科学出版社, 2010: 157-160; 唐英杰, 邹英都. 晚商“畿内田猎区”考论 [J]. 历史研究, 2022 (1).
- [31] 当某一实体类别在语料中出现频率较低时, 模型缺乏足够的语境样本来学习其分布特征, NLP领域中称为长尾分布问题, 参见 Alshammari, et al. Long-Tailed Recognition via Weight Balancing [C] //Proceedings of CVPR. New Orleans: IEEE, 2022: 6897-6907.
- [32] Demsar J. Statistical Comparisons of Classifiers over Multiple Data Sets [J]. Journal of Machine Learning Research, 2006, 7: 1-30.
- [34] 朱凤瀚. 商周家族形态研究 (增订本) [M]. 天津: 天津古籍出版社, 2004: 27-117.
- [35] 董作宾. 卜辞中的亳与商 [C] //董作宾先生全集 (乙编第3册). 台北: 艺文印书馆, 1977: 611; 陈梦家. 殷虚卜辞综述 [M]. 北京: 中华书局, 1988: 255-258; 岛邦男著, 温天河, 李寿林译. 殷墟卜辞研究 [M]. 台北: 鼎文书局, 1975: 689-695; 钟柏生. 殷商卜辞地理理论丛 [M]. 台北: 艺文印书馆, 1989: 39-50; 郑杰祥. 商代地理概论 [M]. 郑州: 中州古籍出版社, 1994: 19-25; 李雪山. 商后期王畿行政区划研究 [J]. 郑州大学学报 (哲学社会科学版), 2001 (2); 孙亚冰, 林欢. 商代地理与方国 [M]. 北京: 中国社会科学出版社, 2010: 139-143.
- [36] 张惟捷. 殷商武丁时期人物“雀”史迹研究 [J]. 历史语言研究所集刊, 2014 (第八十五本第四分).
- [37] 郑杰祥. 商代地理概论 [M]. 郑州: 中州古籍出版社, 1994: 215-217; 钟柏生. 殷商卜辞地理理论丛 [M]. 台北: 艺文印书馆, 1989: 137; 陈青荣. 卜辞危方地望考 [J]. 管子学刊, 2009 (2); 孙亚冰, 林欢. 商代地理与方国 [M]. 北京: 中国社会科学出版社, 2010: 398; 陈絮, 王旭东. 殷墟甲骨中的望乘与望地——兼谈卜辞“虎方”之有无 [J]. 兰州大学学报 (社会科学版), 2019 (6).

(责任编辑: 赵静)