

DP理论在大语言模型产出会话礼貌测评中的应用

——聚焦日语请求行为

西安外国语大学 日本文化经济学院

毋育新

王茹洪

[摘要] 基于宇佐美まゆみのDP理论(话语礼貌理论),聚焦日语拒绝场景的请求行为,比较了自然会话与大语言模型产出会话请求行为的“语步”和“表述连锁链”,旨在考察大语言模型在语用行为建构中的表现,以明确其话语礼貌实现能力。研究发现:两类会话在“追加说明”“感谢”和“插入后话题”的“语步”使用上存在显著性差异;大语言模型产出会话的整体顺序和“表述连锁链”与自然会话一致性程度高,但在模拟少数“语步”时存在局限,特别是缺少了自然会话“基本态”中含有的“顾及”“追加说明”和“插入后话题”三大关键“语步”,脱离了自然会话“基本态”,产生了有标记性行为,引发负向礼貌效果,降低了礼貌度。总体而言,大语言模型虽已具备一定的话语礼貌实现能力,但仍未达到自然会话的礼貌水平。未来还需聚焦话语层面进一步提升其话语礼貌实现能力。

[关键词] 大语言模型 日语请求行为 话语礼貌理论 语步 表述连锁链

引言

2022年美国OpenAI公司产品ChatGPT的正式发布,标志着人工智能从专用型向通用型发展迈出了一大步,为语言学研究提供了新的视野。特别是其突出的语言理解和生成能力引发了语言学界的持续关注和广泛讨论^[1]。人工智能时代背景下,开展语言学与人工智能交叉学科研究成为新的热点。虽然ChatGPT等大语言模型在自然语言处理方面已经取得了巨大的进步,但相较语音、语法、语义来说,语用仍是一大难点。如何从语言学的角度出发,完善大语言模型的自然语言处理逻辑,是一个重大课题。礼貌,作为交流的润滑剂,在人际交往中必不可少,是语用学研究经久不衰的话题。在人工智能时代背景下,有必要从语用学的角度出发,明晰人机对话系统语言生成的特点^[2]。二十一世纪“对话系统建构(Dialogue System Con-

struction)”的持续发展,要求我们要走在时代前列,认定人类与人工智能对话“基本态”,在假定人类与AI共存的基础上,研究礼貌现象的课题^[3]。

1 文献综述

近年来,基于大语言模型的语用学相关研究逐渐兴起,代表性成果包括Hu *et al.* (2023)、刘海涛和亓达(2024)、Barattieri di San Pietro *et al.* (2023)、毛延生和王一航(2024)、Dynel(2023)、山崎等(2022)、松本等(2024)、Andersson & McIntyre (2025)、陈新仁(2024)、李瑶等(2024)、申智奇(2025)。上述研究主要聚焦于以下三个方面。

1.1 大语言模型语用能力研究

聚焦大语言模型语用能力的有Hu *et al.* (2023)、刘海涛和亓达(2024)、Barattieri di San Pietro *et al.* (2023)、毛延生和王一航(2024)、Dynel (2023)。

Hu *et al.* (2023)、刘海涛和亓达(2024)通过欺骗、间接话语、反语等七类语用任务,系统测试了多个大语言模型的语用理解能力,发现大语言模型在多数任务中接近人类水平,对语境线索的敏感度与人类呈显著相关性,但存在处理反语和幽默等语用现象时,依赖于字面意义的直接理解,对复杂语境的推理能力有限的问题,这表明大语言模型缺乏深层社会常识,揭示了当前语言技术在复杂社交情境中的局限性。Barattieri di San Pietro *et al.* (2023)评估了ChatGPT-3.5的语用能力,发现其表达性和接受性语用能力接近人类的表现,但存在违反格赖斯合作原则中的数量准则、语境推理能力(隐喻解读及幽默理解能力)不足等问题。毛延生、王一航(2024)则通过人机共生语境实证研究了ChatGPT-3.5的汉语语用能力,发现其在语用规则明确的道歉场景表现优异,但在请求、安慰、抱怨等言语行为中策略趋同且语境适应力不足,揭示了该模型虽具备表层语言得体性,却难以把握汉语社会心理规范与修辞特征,存在机械翻译痕迹和情感表达程式化问题,强调了本地化训练的必要性。Dyner(2023)基于Reddit's r/ChatGPT社区的互动数据,系统分析了人类与AI对话中呈现的元语言、元交际、元话语及元语用特征,结果表明ChatGPT的回应表征了训练数据中内化的人类语言模式,但存在因指令冲突产生非预期的语用矛盾的问题,反映出AI在模仿人类元语言行为时仍存在伦理边界与技术局限。

1.2 礼貌问题研究

近年来,聚焦大语言模型中礼貌问题的有山崎等(2022)、松本等(2024)、Andersson & McIntyre (2025)。上述三项研究都采用了Brown & Levinson (1987)的礼貌策略理论作为分析框架。山崎等(2022)将人际关系分解为“角色关系”(如上司/下属)和“心理距离”(亲密/普通/疏远)两种,通过在指令中加入人际关系指示,引导模型生成符合特定社交情境的回应,发现模型能有效识别角色关系(如对上司使用高礼貌策略),但对心理距离的细微把控仍存在局限,尤其在非对称关系中对话

自然度较低。松本等(2024)探讨了在条件推理中说话者性格(友好/难相处)与信息提供主体(人类/AI)对逻辑推理抑制的影响,揭示了人机交互中社会认知的差异,为AI对话系统的礼貌策略设计提供了实证依据。Andersson & McIntyre (2025)评估了ChatGPT-3.5识别和解释不礼貌语言的能力,通过测试其在显式(如直接侮辱)、隐式(如讽刺、反问)及元语言情境下的产出表现,发现模型能有效识别常规的不礼貌表达,但在处理依赖复杂语境推理的隐式不礼貌时存在困难,指出尽管ChatGPT展现了一定的语用意识,但其设计依赖训练数据和伦理规则,难以完全模拟人类对社交动态的灵活判断,凸显了AI在复杂语用推理上的局限性。

1.3 大语言模型赋能语用教学研究

聚焦大语言模型赋能语用教学的有陈新仁(2024)、李瑶等(2024)、申智奇(2025)等。陈新仁(2024)基于教学测试实践,表明ChatGPT完全可以赋能二语语用教学,借助多种活动设计,可以在培养学生二语语用语言能力和二语社交语用能力方面发挥重要支撑作用。李瑶等(2024)考察了在汉日语境中,文心一言、ChatGPT、Hugging face三种大语言模型对于语用学中礼貌知识的生成能力,指出虽然大语言模型具备对话流畅性和知识涌现能力,但其生成内容存在准确性不足、完整性缺失和稳定性波动三大缺陷,尤其表现出理论滞后性和自评盲目性等明显不足,表明当前大语言模型尚难胜任语言研究与教学中的专业礼貌知识提供者角色。申智奇(2025)指出数智技术通过真实互动情境、语用特征分析和规范语篇生成赋能语用教学,强调大语言模型在跨文化对比、会话含意解析等方面具有潜力,但也面临技术局限(如对反语和幽默的理解不足)、教师数字素养不足及学术伦理风险等问题。

综上所述,当前基于大语言模型的语用学研究主要集中在语用能力、礼貌问题和语用教学方面,且已积累了较多成果。多项研究共同揭示了大语言模型语用能力的双重性:既具备基于统计规律的表层规则模仿能力,又缺乏深层认知框架

支撑的语用推理机制。主流研究注重通过言语行为考察大语言模型的语用能力,对大语言模型产出的汉语、英语言语行为,特别是请求行为的考察已有所涉及。请求行为作为一种典型的面子威胁性言语行为,其实施过程涉及复杂的语用调控。请求方在交际时需要考虑对方的面子需求缓和其话语力度,还需要综合考虑信息传递、语篇连接等因素^[4]。这一复杂过程不仅可作为衡量个体语用能力的重要维度,亦是探究大语言模型与人类语用表现差异的关键切入点。现有研究中对大语言模型产出的日语言语行为,特别是请求行为的考察还鲜有涉及。同时,聚焦礼貌问题开展的研究多采用 Brown & Levinson(1987)的礼貌策略理论,侧重从单句、单个表述行为层面展开考察,对大语言模型在话语层面礼貌实现程度的考察仍显不足。因此,本研究聚焦言语行为与礼貌问题,从话语层面出发,考察大语言模型在语用行为建构中的表现,旨在明确其话语礼貌实现能力。

2 理论基础

Brown & Levinson(1987)的礼貌策略理论(universal politeness theory)是二十世纪有关礼貌现象研究最具影响力的理论之一,宇佐美まゆみ对 Brown & Levinson(1987)的礼貌策略理论进行了补充与延展,提出了从话语层面研究礼貌问题的 DP 理论(discourse politeness theory),也就是话语礼貌理论,其代表了二十一世纪礼貌现象研究的一个新方向^[5]。Usami(2006)指出,DP 理论将减轻面子威胁的礼貌称为“有标记性礼貌(marked politeness)”,不发生面子威胁但必须采取的约定俗成的礼貌称为“无标记性礼貌(unmarked politeness)”。构成无标记性礼貌的典型场景、状态称为“基本态(discourse politeness default)”。符合“基本态”的行为称为无标记性行为(unmarked behavior),由无标记性行为构成的“基本态”是符合无标记性礼貌的最佳状态。脱离“基本态”的行为称为有标记性行为(marked behavior),有标记性行为会引发相对礼貌效果:负向礼貌效果、中性礼

貌效果或正向礼貌效果。

谢オ(2001)、母育新和郅永玮(2010)、项往等(2022)基于 DP 理论开展的自然会话请求行为研究证实了 DP 理论的科学性和可操作性,日语自然会话的请求行为已得到充分探讨。雒美敏和母育新(2023)将 DP 理论引入人机对话系统研究,验证了 DP 理论在人机对话系统研究中的可行性,拓展了 DP 理论的适用范围。本研究延展雒美敏和母育新(2023)的研究,试将 DP 理论引入大语言模型研究中,拟在 DP 理论的框架下对大语言模型产出日语会话请求行为的建构展开考察,测评其礼貌程度,以明确大语言模型的话语礼貌实现能力。有鉴于此,本研究基于宇佐美まゆみ(2019、2024)的 DP 理论,从无标记性礼貌视角出发,聚焦日语拒绝场景的请求行为,对利用自然会话语料库形式所收集的自然会话语料和以大语言模型间对话形式所收集的 AI 会话语料进行对比。基于“话语基本态”的观点,分别考察日语自然会话与大语言模型产出会话请求行为整体的“典型展开模式”。参考 Swales(1981)学术引言遵循较为固定的结构,这种结构可分为若干“语步(move)”,每个“语步”都有特定的交际目的(communicative purpose)的观点,将请求行为分为若干具有特定交际目的的“语步”,分别考察构成日语自然会话与大语言模型产出会话请求行为的“语步”。参考宇佐美まゆみ(2024)和母育新(2014)的观点,从话语层面出发,将请求行为视为由若干“语步”构成的“表述连锁链(utterance sequence)”所展现的动态过程,分别考察构成日语自然会话与大语言模型产出会话请求行为的“表述连锁链”。基于上述研究内容,拟提出以下两个研究问题:

(1)日语拒绝场景自然会话与大语言模型产出会话请求行为的“语步”是否存在差异?

(2)日语拒绝场景自然会话与大语言模型产出会话请求行为的“表述连锁链”是否存在差异?

3 语料概要

本研究的日语语料,由日语母语者自然会话

语料(JN语料)和大语言模型产出会话语料(AI语料)两部分构成。

JN语料选自宇佐美まゆみ(2023)开发的『BTSJ 1000人日本語自然会話コーパス』(BTSJ 1000人日语自然会话语料库),具体语料为该语料库中“同性好友间的拒绝电话对话(与同年级学生)”部分。该组语料包含13组关系亲密的女性好友之间的电话对话,共计26名参与者(请求方13人,被请求方13人)。调查对象均为以日语为母语的女性大学生或研究生,请求方与被请求方为同年级好友关系。

AI语料参照JN语料的对象及场景设定开展实验获取。实验基于ChatGPT-4o高级语音模式平台,采用两部设备,将一方设定为请求方,另一方设定为被请求方,分别给两方下达指令,指令包含身份、场景、任务和注意事项,命令两方对话,完成拒绝场景中的请求行为,对话过程采用第三部设备进行音频录制。实验指令如表1所示。

语料收集后,依据宇佐美まゆみ(2020)的『基本的な文字化の原則(BTSJ)2019年改訂版』(基本文字化原則2019年修订版),对大语言模型产出会话语料进行文字转写。全部语料详情如表2所示。

表1 大语言模型对话实验指令

实验角色	实验指令
请求方	<p>あなたの名前は高橋智子で、東京大学の大学3年生の女子学生だとします。国立国語研究所で行われている言語調査実験に参加します。実験は明日の午後3時に始まり、3時30分に終わる予定です。ですが、明日、急に用事ができて、あなたは行けなくなりました。友達の中村あゆみに頼んで、自分の代わりに行ってもらいたいと思っています。これからあなたは友達の中村あゆみに電話をかけて依頼を行います。あなたは高橋智子としての身分で、中村あゆみとの関係に基づいて会話を始めますが、会話中の言語表現の論理とポライトネス・ストラテジーに注意する必要があります。準備ができましたら、私からの指示を待ってください。</p> <p>假设你是高桥智子,是东京大学的一名大三女学生。你报名参加了国立国語研究所开展的一项语言调查实验,实验预定明天下午3:00开始,3:30结束。但你临时有急事,明天不能参加,需要请求好友中村步美替自己去参加。现在你要与中村步美通话说明该请求,请基于你的身份,以及你和中村步美的关系发起并展开对话,注意通话过程中的语言表达逻辑和礼貌策略。准备好后,请等待我的指令再开始。</p>
被请求方	<p>あなたの名前は中村あゆみで、東京大学の大学3年生の女子学生だとします。友達の高橋智子から依頼の電話がかかってきました。あなたは中村あゆみとしての身分で、高橋智子との関係に基づいて、情報を聞いて応答し、会話を展開してください。会話中の言語表現の論理とポライトネス・ストラテジーに注意する必要があります。最終的に、あなたは高橋智子からの依頼を断ります。準備ができましたら、高橋智子が会話を始めるのを待ってください。</p> <p>假设你是中村步美,是东京大学的一名大三女学生。你的好友高桥智子与你通话请求你帮她做一件事,你需要基于你的身份,以及你和高桥智子的关系,听取信息并作出回应,展开对话。注意通话过程中的语言表达逻辑和礼貌策略。最后你需要拒绝高桥智子的请求。准备好后,请等待高桥智子发起对话再做出回应。</p>

资料来源:笔者参照宇佐美まゆみ(2023)『BTSJ 1000人日本語自然会話コーパス』的场景设定制作。

注:为便于大语言模型开展对话,实验采用随机生成的姓名为双方设定身份。

表2 语料详情

组别	对话类型	亲疏关系	上下关系	性别	对话数量	对话总时长	平均每组对话时长
JN	好友间的拒绝电话会话	亲密	同年级	女	13组	26分39秒	2分3秒
AI					13组	15分49秒	1分13秒

资料来源:笔者依据宇佐美まゆみ(2023)『BTSJ 1000人日本語自然会話コーパス』及实验调查语料制作。

注:尽管本研究的语料数量有限,但JN语料由日本国立国語研究所宇佐美教授收集并严格筛选,具有较好代表性。同时,AI语料参照JN语料的标准筛选,因此在一定程度上也具备较高代表性。

4 日语请求行为的人机对比分析

4.1 日语请求行为“语步”的编码

借鉴项往等(2022)的研究,从日语自然会话语料中析出如表3所示的14个“语步”。使用表3的14个“语步”,分别对各13组JN语料、AI语料进行编码,不符合表3“语步”的表述行为标注为“无”。编码过程如下页表4所示。

为确保“语步”标注的准确性,本研究设立第三判定者,使用Cohen's Kappa系数分析对两位判定者的“语步”标注结果进行一致性检验。Cohen's Kappa系数为0.78($p < 0.01$),表明两名判定者的判断具有较高一致性,数据分析结果可靠。

4.2 日语请求行为“语步”的人机对比

为明确日语拒绝场景自然会话与大语言模型产出会话请求行为的“语步”是否存在差异,本研

究对日语自然会话与大语言模型产出会话请求行为的“语步”的使用频率进行统计,并对每对“语步”的使用频率进行比较。

首先,统计JN和AI中各个“语步”的使用频率。“使用频率”是指在对话中是否包含该“语步”,即使某一“语步”在同一对话中出现多次,也仅记为一次。利用Python pandas库,遍经过“语步”编码的JN和AI的Excel语料文件,提取请求说话者的“语步”使用频率及其具体所在文件、位置,并生成包含统计结果的“语步”统计Excel文件。

其次,再次使用Python pandas库读取生成的“语步”统计Excel文件,提取JN和AI的“语步”使用频率。针对每对“语步”,构建2×2列联表,如下页表5为表示“会话类型(JN和AI)”和“开场白语步(有/无)”这两个分类变量间关系的2×2列联表。计数资料的组间比较通常采用卡方检验,当样本量(N)≥40且理论频数(T)≥5时,可以直接采用卡方检

表3 构成JN请求行为的“语步”

序号	语步	定义	例句(JN语料)
1	引起注意	电话中与受话者打招呼的话语。	もしもし。(喂。)
2	顾及	确认受话者是否方便交谈的话语。	なんか、大丈夫?(那个,现在方便吗?)
3	确认日程	确认受话者日程安排的话语。	明日なんかテストとかある?(明天有考试之类的安排吗?)
4	插入前话题	施行请求行为前插入与请求不相关的话题。	あ、あたしもみたみた、さっきのテレビ。(我也看了,刚才的电视节目。)
5	开场白	预告请求行为的话语。	ちょっと用事があるんだけど。(我有事想麻烦你。)
6	情况说明	说明事情具体情况的话语。	国立国語研究所っていうところでね、日本語の会話調査に関する実験っていうのがあって。(国立国語研究所有个关于日语会话调查的实验。)
7	请求	施行请求行为的话语。	もし時間があつたら、代わりに行ってくれないかなっていうお願いなんだけど。(如果有时间的话,能不能帮我去一趟呢?)
8	追加说明	补充请求内容相关细节的话语。	月曜日の朝9時から、ちょっと遠いところなんだけど。(星期一早上9点,有点远。)
9	再次请求	施行请求行为后再次发起请求的话语。	だから、もし、時間あれば朝のね、9時からなんだけど。(所以,如果早上有时间的话,9点开始。)
10	替代方案	提出请求的替代方案的话语。	あ、じゃほかの人を探してみるよ。(啊,那我找找其他人吧。)
11	道歉	施行请求行为后,表达歉意的话语。	ごめんね。(对不起。)
12	感谢	施行请求行为后,表达感谢的话语。	ありがとう。(谢谢。)
13	插入后话题	施行请求行为后插入与请求不相关的话题。	明日部活行く?(你明天去社团活动吗?)
14	结束对话	结束电话对话的话语。	じゃまた。(拜拜。)

资料来源:笔者依据项往等(2022)制作,有部分修改。

表4 语料编码过程例

说话者	表述内容	语步标注
AI-JFB013	もしもし、神崎雅美です。(你好,我是神崎雅美。)	引起注意
AI-JFB013	急にごめんね。(突然打扰,实在抱歉。)	顾及
AI-JFB013	明日の午後、国立国語研究所で行われる言語調査実験に参加する予定だったんだけど、急用が入っちゃって行けなくなっちゃったの。(我本要参加明天下午国立国語研究所开展的语言调查实验,但临时有急事,去不了了。)	情况说明
AI-JFB013	もし時間があれば代わりに行ってもらえないかな?(如果你有空的话,能代替我去吗?)	请求
AI-JFSa013	雅美、こんにちは。(你好,雅美。)	
AI-JFSa013	急なことで大変だね。(这么突然,真不好办呢。)	
AI-JFSa013	明日の午後、ちょっと予定があつて難しいかもしれないんだけど、力になれなくてごめんね。(明天下午,我有点安排,可能有点困难,不能帮到你真抱歉。)	
AI-JFB013	わかりました。(我明白了。)	无
AI-JFB013	気にしないで、ありがとう。(没关系,还是谢谢你。)	感谢
AI-JFB013	また今度ね。(下次再聊。)	结束对话
AI-JFSa013	うん、また今度ね。(嗯,下次再聊。)	
AI-JFSa013	何かあったらいつでも言ってね。(要是还有什么事随时联系我哦。)	
AI-JFB013	うん、また話そうね。(嗯,下次再聊呢。)	无
AI-JFB013	バイバイ。(拜拜。)	结束对话

资料来源:笔者依据基于宇佐美まゆみ(2020)的『基本的な文字化の原則(BTSJ)2019年改訂版』转写原则及项往等(2022)“语步”认定标准,对语料的编码标注结果制作。

注:JFB表示请求方,JFSa表示拒绝方(被请求方)。

表5 会话类型与“开场白”语步使用情况间的2×2列联表

会话类型 \ “开场白”语步	“开场白”语步		总计
	有	无	
JN	9(69.2%)	4(30.8%)	13
AI	11(84.6%)	2(15.4%)	13
总计	20	6	26

资料来源:笔者依据 Python pandas 库统计结果制作。

验;但当N<40或T<1时,则应选择 Fisher 精确概率法^[6]。鉴于各“语步”的2×2列联表中,样本量(N)均小于40,直接采用卡方检验不符合适用条件。因此,本研究采用 Fisher 精确概率法,根据构建的2×2列联表,借助 Python scipy.stats 库中的 fisher_exact 函数对 JN 和 AI 的全部“语步”使用情况进行 Fisher 精确检验,以判断 JN 和 AI 的“语步”使用情

况是否存在显著性差异。检验结果如下页表6所示。

依据表6 Fisher 精确检验结果可知,在“追加说明”和“感谢”语步的使用方面,JN 和 AI 在5%的显著性水平上存在显著性差异,在“插入后话题”的使用方面,JN 和 AI 在1%的显著性水平上存在显著性差异。JN 使用“追加说明”和“插入后话题”的频率高,而 AI 使用“感谢”的频率高。

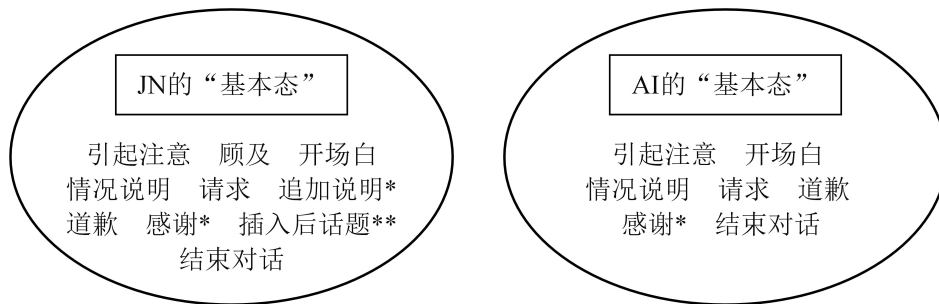
最后,根据宇佐美まゆみ(2001、2008)提出的“基本态”认定方法,将使用频率超过50%的“语步”定义为“基本态”。笔者将构成 JN 和 AI 请求行为的“语步”的“基本态”归纳总结于下页图1。对比发现,“顾及”“追加说明”和“插入后话题”仅出现在 JN 请求行为的“基本态”中,而未出现在 AI 请求行为的“基本态”中。

表6 各“语步”使用次数及 Fisher精确检验结果

序号	语步	JN对话使用次数(%)	AI对话使用次数(%)	有无显著性差异(p)
1	引起注意	13(100.0%)	13(100.0%)	无(p>0.05)
2	顾及	9(69.2%)	6(46.2%)	无(p>0.05)
3	确认日程	4(30.8%)	0(0.0%)	无(p>0.05)
4	插入前话题	1(7.7%)	1(7.7%)	无(p>0.05)
5	开场白	9(69.2%)	11(84.6%)	无(p>0.05)
6	情况说明	13(100.0%)	12(92.3%)	无(p>0.05)
7	请求	13(100.0%)	13(100.0%)	无(p>0.05)
8	追加说明	9(69.2%)	3(23.1%)	有*(p<0.05)
9	再次请求	5(38.5%)	1(7.7%)	无(p>0.05)
10	替代方案	4(30.8%)	6(46.2%)	无(p>0.05)
11	道歉	11(84.6%)	7(53.8%)	无(p>0.05)
12	感谢	8(61.5%)	13(100.0%)	有*(p<0.05)
13	插入后话题	9(69.2%)	1(7.7%)	有**(p<0.01)
14	结束对话	12(92.3%)	12(92.3%)	无(p>0.05)

资料来源:笔者依据Python pandas库统计结果及scipy.stats库fisher_exact函数的Fisher精确检验结果制作。

注:*p<0.05、**p<0.01。



资料来源:笔者依据实验调查统计分析结果制作。

注:*p<0.05、**p<0.01。

图1 JN和AI请求行为“语步”的“基本态”

4.3 日语请求行为“表述连锁链”的人机对比

对于“表述连锁链”的认定,项往等(2022)参考謝オン(2001)的研究成果,将“→”用于表示会话的进行方向,“+”用于连接因个体差异而顺序可互换的“语步”,明确了汉日拒绝场景请求行为的“表述连锁链”。笔者在此基础上,借助Python的数据处理与科学计算工具库进行科学计算,拟进一步相对客观地归纳具有较高代表性,能够作为

“基本态”的“表述连锁链”。

首先,分析JN和AI的整体顺序。使用Python pandas库遍历JN和AI的每个Excel语料编码文件中请求方的“语步”列,记录每个“语步”的文件来源、行号。由于每个语料文件的对话长度(即行数)不同,无法直接比较每个“语步”的绝对位置(即行号)。因此,需要将“语步”的绝对位置转换为归一化位置(即相对位置),计算公式如下:

$$\text{归一化位置} = \frac{\text{语步所在行号}}{\text{文件总行数}}$$

归一化(normalization)是一种简化计算的方式,即将所有属性以相同的测量单位表示,并使用通用的刻度或范围^[7]。García *et al.* (2015)和Abram *et al.* (2022)认为不同评价指标往往具有不同量纲,数值间差别较大,不进行处理可能会影响到数据分析的结果。为消除指标之间的量纲和取值范围差异的影响,需要对输入数据进行归一化处理。归一化试图赋予所有数据属性同等的权重,使属性之间的比较与聚合更容易,数据的收敛条件更好。Tang *et al.* (2021)指出,归一化即将数据按照比例进行缩放,落入一个特定的区域,以便进行综合分析。本研究采用归一化方法,对Excel语料编码文件中全部“语步”位置值进行统一缩放,以确保不同语料中“语步”位置的可比性与等权性。具体来说,如果一个语料编码文件总共有100行,而某个“语步”出现在第10行,则该“语步”的归一化位置为10/100=0.10。通过这种方式,所有“语步”的位置都被转化为0到1的范围内,以便进行对比。利用Python numpy库计算每个“语步”

的归一化位置,为避免极值影响,采用每个“语步”归一化位置的中位数(如表7所示),作为排序依据。

为了考察大语言模型产出会话(AI)的整体顺序与自然会话(JN)是否一致,本研究采用斯皮尔曼(spearman)相关分析、Bland-Altman分析和组内相关系数(ICC)分析,从不同角度评估两组数据的一致性。检验结果分别如表8、下页图2、下页表9所示。

表8 “语步”归一化位置中位数的斯皮尔曼相关分析

变量对	相关系数(r _s)	显著性水平(p)	样本量(N)
JN与AI“语步”归一化位置中位数	0.88	p<0.01	14

资料来源:笔者依据SPSS软件检验结果制作。

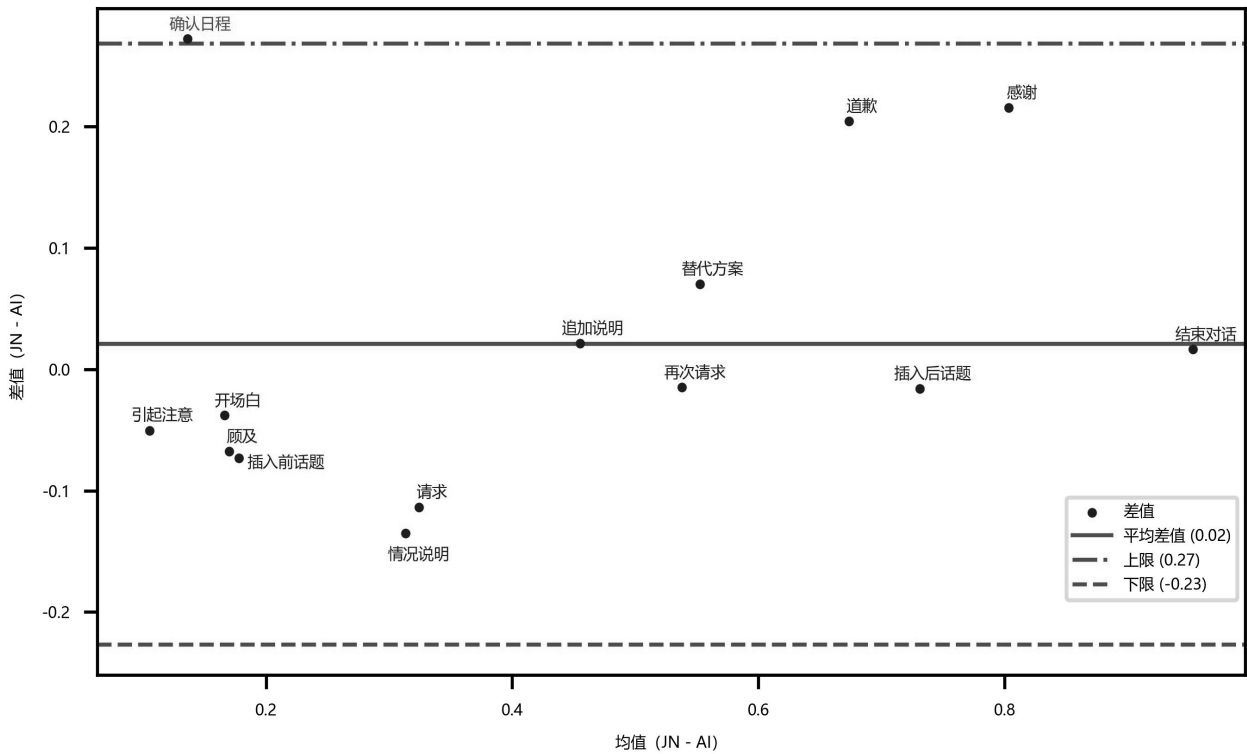
斯皮尔曼(spearman)相关分析结果显示,AI和JN的“语步”排序具有显著正相关(r_s=0.88, p<0.01),表明AI在模拟JN的“语步”顺序特征方面表现出较强的能力。

Bland-Altman分析进一步验证了两组数据在

表7 “语步”归一化位置中位数

语步	JN语料出现次数	AI语料出现次数	JN归一化位置中位数	AI归一化位置中位数
引起注意	16	13	0.080128205	0.130434783
顾及	15	6	0.136363636	0.203933747
确认日程	4	0	0.272427078	0
插入前话题	5	2	0.141304348	0.214285714
开场白	9	11	0.147540984	0.185185185
情况说明	39	14	0.245901639	0.380952381
请求	16	13	0.267526777	0.380952381
追加说明	55	3	0.465753425	0.444444444
再次请求	6	1	0.530598571	0.545454545
替代方案	4	7	0.587662338	0.517241379
道歉	20	7	0.775919732	0.571428571
感谢	11	27	0.911392405	0.695652174
插入后话题	26	1	0.723404255	0.739130435
结束对话	25	20	0.961538462	0.944940476

资料来源:笔者依据Python pandas库统计结果及numpy库计算结果制作。



资料来源:笔者依据 Python 检验结果制作。

图2 Bland-Altman分析图

表9 ICC分析结果

类型	描述	ICC	F	置信区间(95%)	显著性水平(p)
ICC2	单次测量的一致性	0.906	19.418	[0.74, 0.97]	p<0.01
ICC2k	多次测量的平均一致性	0.951	19.418	[0.85, 0.98]	p<0.01

资料来源:笔者依据 Python 检验结果制作。

绝对值上的一致性,绝大多数数据点位于95%的一致性限度([-0.23, 0.27])内,且差值均值(0.02, SD=0.13)接近于0,表明AI和JN的整体系统性偏差较小。然而,如图2和表10所示,由于“确认日程”语步未出现在AI中,导致其差值超出一致性范围。这一异常点揭示了大语言模型在对低频或上下文依赖性较强的“语步”模拟时仍存在局限。

表10 Bland-Altman分析异常点

异常点	语步	均值	差值
上限外	确认日程	0.136214	0.272427

资料来源:笔者依据 Python 检验结果制作。

组内相关系数(ICC)分析结果显示,单次测量一致性(ICC2=0.906)和多次测量平均一致性(ICC2k=0.951)均达到较高一致性水平,进一步验证了AI与JN的顺序具有高度一致性。

综合三种分析结果可以得出结论:大语言模型产出会话在大多数“语步”上能够较好地模拟自然会话的整体顺序特征,但仍需改进对特定“语步”(如“确认日程”)的捕捉能力,以进一步提升一致性和可靠性。

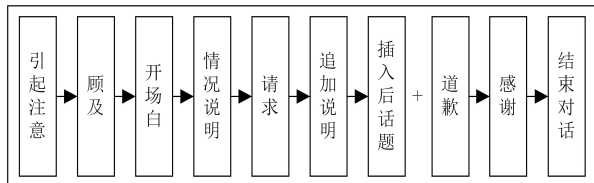
其次,依据会话的整体顺序,归纳分析能够作为请求行为“基本态”的“表述连锁链”。筛选使用频率超过50%的“语步”,按归一化位置中位数升序排列,生成初步的“表述连锁链”。为明确顺

序可互换的“语步”，使用 Python itertools 库生成所有 Excel 语料编码文件中“语步对”的组合，统计每对“语步”(A, B)的相对顺序。例如，统计“语步 A”出现在“语步 B”之前的次数(记作 *AbeforeB*)和“语步 B”出现在“语步 A”之前的次数(记作 *BbeforeA*)。使用以下公式计算相对顺序比例：

$$\text{相对顺序比例} = \frac{A_{\text{before}B}}{A_{\text{before}B} + B_{\text{before}A}}$$

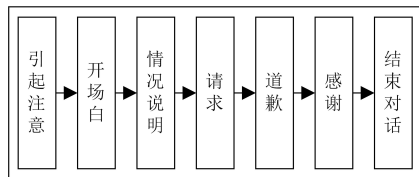
若每对“语步”的相对顺序比例接近 50%(例如在 40~60%范围内)，则认为该“语步对”的顺序可以灵活调换。

最后，根据初步的“表述连锁链”和“语步”相对顺序比例，生成最终具有较高客观代表性，能够作为“基本态”的“表述连锁链”。JN 请求行为的“表述连锁链”如图 3 所示，AI 请求行为的“表述连锁链”如图 4 所示。



资料来源：笔者依据调查统计分析结果制作。

图3 JN 请求行为的“表述连锁链”



资料来源：笔者依据调查统计分析结果制作。

图4 AI 请求行为的“表述连锁链”

4.4 日语请求行为礼貌程度的人机对比

如上所述，日语拒绝场景自然会话与大语言模型产出会话的请求行为在“语步”和“表述连锁链”上均存在一定差异。

JN 和 AI 的请求行为由“请求前准备”“施行请求行为”和“请求后应对”三个部分构成。

在请求前准备阶段，JN 通过“引起注意”→

“顾及”→“开场白”→“情况说明”的表述序列，渐进切入话题，以缓和面子威胁行为，为施行请求行为做铺垫。其中“顾及”作为“基本态”语步，通过预设听话者立场(例 JFB003: あのね、今電話大丈夫? [那个，现在方便通话吗?])降低对话的强加性。而 AI 缺失“顾及”语步，直接以“开场白”发起对话(例 AI-JFB010: ちょっとお願いしたいことがあるんだけど、いいかな? [我有件事想拜托你，可以吗?])(下页表 11 第 2 行)，未预先照顾听话者的消极面子需求，脱离了 JN“基本态”。

在施行请求行为阶段，AI 与 JN 一致，均使用“请求”(例 AI-JFB010: 代わりに行ってもらえないかな? [能替我去一下吗?])(表 11 第 7 行)；例 JFB003: 行っていただけませんかという。[能帮我去一下吗?])施行请求行为，以完成核心交际目标，符合 JN“基本态”。这一阶段的行为趋同性表明，AI 在显性言语行为的程式化生成上已达到与 JN 相近的水平。

在施行请求行为后的应对阶段，JN 通过“追加说明”→“插入后话题”+“道歉”→“感谢”→“结束对话”的表述序列实现请求行为的闭环。施行请求行为后，JN 通过使用“追加说明”(例 JFB004: うん、金曜日。[嗯，星期五。])补充与请求相关的细节，一方面增加了请求被同意的可能，另一方面也为听话者提供了考虑的时间。在请求被拒绝后，JN 通过使用“插入后话题”(例 JFB010: 明日部活行く? [明天去社团活动吗?])转移对话焦点，缓解因听话者拒绝请求可能引发的交际障碍，维护了与听话者的人际关系。而 AI 缺失作为 JN“基本态”的“追加说明”和“插入后话题”，缺少对对话的动态建构和对人际关系的维护。

综上所述，在日语拒绝场景的请求行为中，大语言模型产出会话由于缺少“顾及”“追加说明”“插入后话题”三大“语步”，缺少了自然会话中体现的渐进式推进、对话动态建构和人际关系维护特征，脱离了自然会话“基本态”。根据 DP 理论，脱离“基本态”会产生有标记性行为，而有标记性行为的相对礼貌效果取决于说话者和听话者间的“估算差”^[8]。而就此处来看，作为人类的听话者期

表 11 大语言模型产出会话例

行号码	说话者	表述内容	语步标注
1	AI-JFB010	もしもし佐々木雪絵、彩香です。	引起注意
2	AI-JFB010	ちょっとお願いしたいことがあるんだけど、いいかな?	开场白
3	AI-JFSa010	彩香、久しぶり。	
4	AI-JFSa010	どうしたの?	
5	AI-JFSa010	何か手伝えることがあれば教えて。	
6	AI-JFB010	ありがとう、雪絵。	无
7	AI-JFB010	実は明日国立国語研究所の言語調査実験に参加する予定だったんだけど、急な用事が入っていけなくなっちゃって、代わりに行ってもらえないかな?	情况说明;请求
8	AI-JFSa010	彩香大変そうね。	
9	AI-JFSa010	でも、明日は私も予定があって、代わりに行けそうにないの。	
10	AI-JFSa010	ごめんね。	
11	AI-JFB010	そっか、雪絵も予定があるんだね。	无
12	AI-JFB010	無理言ってごめんね。	道歉
13	AI-JFSa010	本当にごめん。	
14	AI-JFSa010	何か別の方法で手伝えることがあれば言ってね。	
15	AI-JFB010	ありがとう、雪絵。	感谢
16	AI-JFB010	別の機会にまたお願いするかもしれないけど、その時はよろしくね。	无
17	AI-JFSa010	もちろん。	
18	AI-JFSa010	いつでも連絡してね。	
19	AI-JFSa010	応援してるよ。	
20	AI-JFB010	本当にありがとう。	感谢
21	AI-JFB010	心強いよ。	无
22	AI-JFB010	またね。	结束对话
23	AI-JFSa010	どういたしまして。	
24	AI-JFSa010	いつでも頼ってね。	
25	AI-JFB010	またよろしくね。	无
26	AI-JFB010	ありがとう。	感谢

资料来源:笔者基于大语言模型对话实验收集语料转写编码后的内容制作。

注:JFB表示请求方,JFSa表示拒绝方(被请求方)。为提高可读性,请求方话语采用灰色底纹突出显示。

待有上述三个“语步”，而AI并未提供，故而估算差值在“-1≤De<-α”间，为“过小行为”，引发负向礼貌效果，造成不礼貌（详细请参见《日语学习与研究》2019年第5期第26页图1），降低了会话的礼貌度，使其未能达到与自然会话相持平的礼貌水平。

5 结 语

本研究将宇佐美まゆみのDP理论引入大语言模型研究中，在DP理论的框架下，从无标记性礼貌视角出发，聚焦言语行为与礼貌问题，从话语层面对比了日语自然会话与大语言模型产出会话请求行为的“语步”和“表述连锁链”，旨在考察大语言模型在语用行为建构中的表现，以明确其话语礼貌实现能力。研究结果表明：(1)大语言模型产出会话与自然会话在“追加说明”“感谢”和“插入后话题”的“语步”使用上存在显著性差异。自然会话使用“追加说明”和“插入后话题”的频率高，而大语言模型产出会话使用“感谢”的频率高；(2)大语言模型产出会话的整体顺序与自然会话一致性程度高，但在模拟低频或依赖上下文的“语步”时存在局限；(3)大语言模型产出会话的“表述连锁链”与自然会话大体上一致，但缺少了自然会话“基本态”中含有的“顾及”“追加说明”和“插入后话题”三大关键“语步”，在请求前准备阶段缺少渐进式推进过程，在请求后应对阶段缺少对话动态建构和人际关系维护过程，脱离了自然会话“基本态”，产生了有标记性行为，引发负向礼貌效果，

降低了会话的礼貌度。总体而言，大语言模型虽已具备一定的话语礼貌实现能力，但其产出会话仍缺失构建“基本态”所需的关键“语步”，在对话的动态语用调控方面表现出不足，仍未达到自然会话的礼貌水平。未来还需聚焦话语层面进一步提升其话语礼貌实现能力。

此外，本研究从话语层面测评大语言模型产出日语会话请求行为的礼貌度，为基于大语言模型的语用学研究提供了新的视角。其次，研究证实了DP理论在大语言模型研究中的适用性和可操作性，拓展了DP理论的应用范畴。最后，研究明确了大语言模型产出会话请求行为产生不礼貌印象的关键，在于其脱离了自然会话请求行为的“基本态”。这一发现为今后大语言模型优化动态语用调控，提升话语礼貌实现能力，改进会话产出提供了新的语用学参考路径。由于篇幅所限，本研究仅考察了日语拒绝场景自然会话与大语言模型产出会话的请求行为，汉语拒绝场景自然会话与大语言模型产出会话的请求行为是否存在差异，大语言模型在汉语语境下的话语礼貌实现能力达到何种程度，拟待今后探讨。

[本文为2021年度国家社会科学基金西部项目“面向人机对话系统的日语语用距离调节机制研究”(批准号:21XYY009)阶段研究成果。项目主持人:毋育新;2025年度西安外国语大学研究生科研基金项目“话语礼貌理论视域下大语言模型产出日语请求行为会话的礼貌度测评研究”(项目编号:2025SS010)阶段研究成果。项目主持人:王茄洪]

注

- [1] 李瑶,于富喜,毋育新.汉日语境中大语言模型的礼貌知识生成能力探析[J].日语学习与研究,2024(5):113.
- [2] 雒美敏,毋育新.中日人机对话系统语言产出对比研究——以网络购物场景为例[J].日语学习与研究,2023(5):102.
- [3] 宇佐美まゆみ.21世纪礼貌现象研究的可能性——话语礼貌理论的新发展[J].日语学习与研究,2019(5):27.
- [4] 项往,储天阳,毋育新.汉日拒绝场景请求言语行为的元话语对比分析[J].外语与外语教学,2024(2):75.
- [5] 毋育新.现代日语礼貌现象研究[M].杭州:浙江工商大学出版社,2014:30-31.
- [6] 武松.SPSS实战与统计思维[M].北京:清华大学出版社,2019:186-188.
- [7] 杨寒雨,赵晓永,王磊.数据归一化方法综述[J].计算机工程与应用,2023(3):14.
- [8] 宇佐美まゆみ.ポライトネス理論:発話行為から談話へ[M].東京:大修館書店,2024:200.

参考文献

- 毛延生, 王一航. 生成式预训练语言模型汉语语用能力的实证研究——以ChatGPT为例[J]. 外国语言文学, 2024(6).
- 毋育新, 鄧永玮. 基于话语礼貌理论的日语请求行为研究[J]. 外语教学, 2010(4).
- 申智奇. 教育数智化、大语言模型与语用能力教学[J]. 赣南师范大学学报, 2025(1).
- 刘海涛, 亓达. 大语言模型的语用能力探索——从整体评估到反语分析[J]. 现代外语, 2024(4).
- 陈新仁. ChatGPT与二语语用教学[J]. 外语教学理论与实践, 2024(4).
- 项往, 毋育新, 储天阳. 话语礼貌理论视域下汉日拒绝场景请求行为对比研究——基于同性之间亲近关系的维度[J]. 日语学习与研究, 2022(3).
- 山崎天, 川本稔己, 吉川克正, 等. 大規模汎用言語モデルを用いた雑談対話システムの対人関係性に基づく発話制御の検討[C]//言語処理学会第28回年次大会発表論文集. 京都: 言語処理学会事務局, 2022.
- 宇佐美まゆみ. 談話のポライトネス: ポライトネスの談話理論構想[M]//国立国語研究所. 談話のポライトネス Discourse politeness. 東京: 国立国語研究所, 2001.
- 宇佐美まゆみ. ポライトネス理論研究のフロンティア——ポライトネス理論研究の課題とディスコース・ポライトネス理論[J]. 社会言語科学, 2008, 11(1).
- 宇佐美まゆみ. 基本的な文字化の原則(BTSJ)2019年改訂版[M]//宇佐美まゆみ. 自然会話分析への語用論的アプローチ: BTSJコーパスを利用して. 東京: ひつじ書房, 2020.
- 宇佐美まゆみ. BTSJ 1000人日本語自然会話コーパス[DB/OL]. 東京: 国立国語研究所, 2023[2025-04-10]. <https://mmsrv.ninjal.ac.jp/btsj/>.
- 松本和紀, 高橋達二, 中村紘子. 条件推論におけるAIの発話の解釈と推論抑制: ポライトネス理論に基づく検討[C]//2024年度日本認知科学会第41回大会発表論文集, 東京: 日本認知科学会, 2024.
- 謝オン. 談話レベルからみた「依頼発話」の切り出し方——日本人大学生同士と中国人大学生同士の依頼発話から[J]. 日本研究教育年報, 2001, 5.
- Abram K J, McCloskey D. A comprehensive evaluation of metabolomics data preprocessing methods for deep learning[J]. Metabolites, 2022, 12(3).
- Andersson M, McIntyre D. Can ChatGPT recognize impoliteness? An exploratory study of the pragmatic awareness of a large language model[J]. Journal of Pragmatics, 2025, 239.
- Barattieri di San Pietro C, Frau F, Mangiaterra V, et al. The pragmatic profile of ChatGPT: Assessing the communicative skills of a conversational agent[J]. Sistemi intelligenti, 2023, 35(2).
- Brown P, Levinson S C. Politeness: Some universals in language usage[M]. Cambridge: Cambridge University Press, 1987.
- Dynel M. Lessons in linguistics with ChatGPT: Metapragmatics, metacommunication, metadiscourse and metalanguage in human-AI interactions[J]. Language & Communication, 2023, 93.
- García S, Luengo J, Herrera F. Data preprocessing in data mining[M]. Cham, Switzerland: Springer International Publishing, 2015.
- Hu J, Floyd S, Jouravlev O, et al. A fine-grained comparison of pragmatic language understanding in humans and language models[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Toronto: Association for Computational Linguistics, 2023.
- Swales J M. Aspects of Article Introductions[M]. Birmingham: University of Aston Press, 1981.
- Tang C, Xu Y, Zhu Q. Data normalization improves semantic annotation - a case study of rare disease name annotation[C]//2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Houston: IEEE, 2021.

Usami M. Discourse politeness theory and second language acquisition[M]//Foreign language teaching in Asia and beyond: Current perspectives and future directions. Singapore: Centre for Language Studies, National University of Singapore, 2006.

作者简介: 毋育新 男 汉族 西安外国语大学日本文化经济学院教授 研究方向: 日语语言

联系方式: E-mail: wuyuxin@xisu.edu.cn

作者简介: 王茄洪 男 汉族 西安外国语大学日本文化经济学院日语语言文学专业硕士研究生 研究方向: 日语语言

联系方式: E-mail: 2209314599@qq.com

Application of Discourse Politeness Theory in the Politeness Assessment of LLM-Generated Conversations: A Focus on Japanese Request Behavior

Abstract: Based on the discourse politeness theory proposed by Usami Mayumi, this study focuses on the request behavior in Japanese refusal scenarios, comparing the "moves" and "utterance sequences" of request behavior in natural conversations with those generated by a Large Language Model (ChatGPT-4o), with the aim of examining the model's performance in constructing pragmatic behavior and realizing discourse politeness. The analysis reveals significant differences in the use of three specific "moves": "additional explanation", "gratitude", and "post-topic insertion". Although the overall structure and "utterance sequence" of LLM-generated conversations are largely consistent with those of natural conversations, the model shows limitations in replicating certain key "moves". In particular, the absence of three core "moves"—"consideration", "additional explanation", and "post-topic insertion"—results in a deviation from the "discourse politeness default" of natural conversations, leading to marked behavior that gives rise to a negative politeness effect and, consequently, a lower overall level of politeness. In summary, while the model demonstrates a certain degree of discourse politeness competence, it still falls short of achieving the politeness level of natural conversations. Further improvement is needed, especially at the discourse level.

Keywords: Large Language Models; Japanese request behavior; discourse politeness theory; move; utterance sequence

Authors' Information:

Wu Yuxin (Male)

Professor at Xi'an International Studies University, China

Japanese Linguistics

E-mail: wuyuxin@xisu.edu.cn

Wang Jiahong (Male)

Master's Student at Xi'an International Studies University, China

Japanese Linguistics

E-mail: 2209314599@qq.com

(责任编辑: 王晨萌)