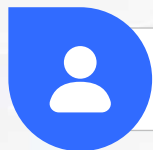


...

基于可接受性的机器空间语义 理解能力评测研究



汇报人：肖力铭



汇报时间：20250917



目录

01

引言

02

影响空间语义可接受性的认知因素

03

面向空间信息正误判断任务的
语义标签设计

04

大语言模型的评测结果分析

05

总结与讨论

一、引言

长期以来，语言学家借助可接受性判断（acceptability judgment）探索人类语言知识的边界，揭示自然语言的规律。

随着人工智能的发展，自然语言处理领域愈发重视对语言模型能力的评估，研究者开始将语言可接受性判断引入机器评测，并逐渐发展为一项基础性任务。

(1) a. 她的弟弟讨厌他自己。

b. *她的弟弟讨厌她自己。 (Liu et al., 2024)

一、引言

现有的可接受性判断数据集可分为三类：

1. 不能说的句子：句法错误、词义搭配错误

(2) a. 那边站着八位舞者。 b. *那边站着八条舞者。 (Liu et al., 2024)

2. 不合理的句子：常识性错误

(3) a. *John put a turkey into the fridge.*
b. **John put an elephant into the fridge.* (Wang et al., 2020)

3. 不通顺的篇章

(4) a. *The lone ranger jumped on his horse. Then he rode into the sunset.*
b. **The lone ranger rode off into the sunset. Then he jumped on his horse.*
(Beyer et al., 2020)

一、引言

现有可接受性判断数据集的特点：

1. 集中关注句法问题，目前已构建多个大规模、综合性的句法可接受性判断数据集。

└─→ 可进一步构建综合性的语义可接受性判断数据集

2. 测试题以包含考察点的单句对为主。

└─→ 不贴近模型在预训练与应用中的真实语言环境

研究目标：

构建一个以文本为测试单位的综合性语义可接受性判断数据集，覆盖词义选择限制、句义合理性、语篇逻辑连贯性等不同层面的问题。

一、引言

语义现象繁多且复杂，有必要先从一个具体的语义范畴切入，探索构建综合型语义可接受性判断数据集的可行路径。

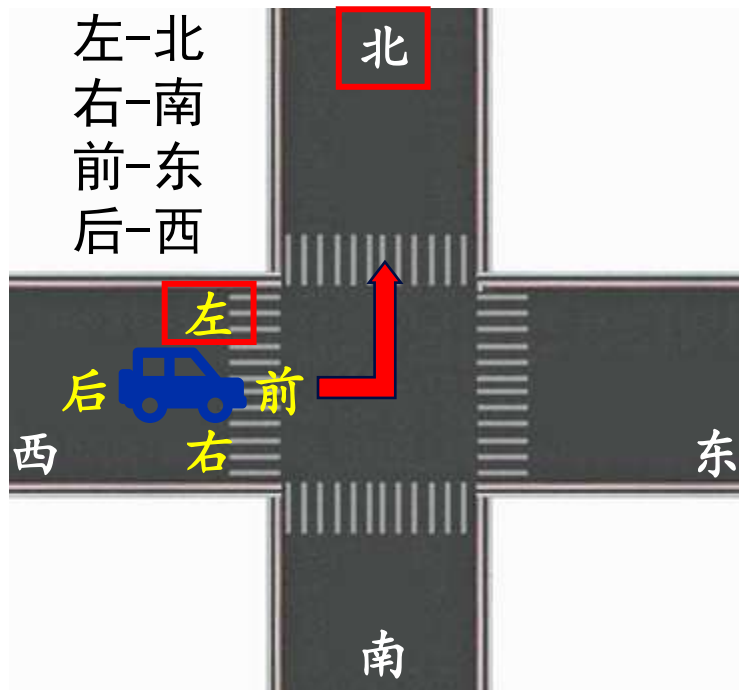
空间语义是一个合适的起点。

1. 物理世界是空间语义理解的认知基础，语义可接受性判断能有较明确的结果。
2. 空间语义理解天然依赖常识和语境，便于考察不同层面的语义可接受性问题。
3. 空间语义理解关联语言世界和现实世界，本身具有较高的认知复杂性，构建的数据集也将具备一定的挑战性。

一、引言

空间信息：位置、起点、终点、轨迹、方向、目标、朝向、部位、形态、距离。

SpaCE系列空间语义理解评测提出了“空间信息正误判断”任务，要求机器判断一个文本的空间信息是否正确或合理。



(5) a. 鄂某某驾驶小汽车沿本市杨浦区控江路由西向东至敦化路路口，遇绿灯向北左转弯至北侧人行横道，碰撞到沿人行横道由西向东行走的被害人张某某。

b. *……由西向东至敦化路路口，遇绿灯向北右转弯至北侧人行横道，……

一、引言

基于“空间信息正误判断”任务分析机器语义理解能力的已有研究存在以下不足之处：

1. 空间语义异常类型“搭配不当、上下文语义冲突、不符合常识或背景信息”有待细化；
2. 定量分析主要基于词性这一形式标签进行，缺少语义维度标签。

研究目标：

细化空间语义异常类型，设计语义维度标签，分析大语言模型的空间语义理解能力。

二、影响空间语义可接受性的认知因素

语义的不可接受不仅表现为语义冲突，在更深层次的认知层面上，反映了认知图式在激活与整合过程中的失败（崔希亮，2001；2002）。

空间语义理解的本质在于将目标物、参照物、方位词等空间要素的认知图式加以整合，在大脑中建构出一个合理的空间场景（spatial scene）。

基于空间场景，空间语义的可接受性判断围绕以下三个问题递进展开：
（1）空间场景能否顺利构建；（2）空间场景是否偏离常规；（3）多个空间场景之间是否存在逻辑矛盾。

二、影响空间语义可接受性的认知因素

本文定义空间场景包括：场域、目标物、事件、观察者

场域：空间发生的地方，由空间介词和参照域组合表达。

(1) 参照域

包括由参照物与方位词组成的方位短语，如“床上”指示床的表面；处所词，如“教室”、“桌面”；由空间介词标记的普通实体，如“沿红色地毯”的“地毯”；“这、那”等空间指示词。

处所词的理解与听话人的生活环境、方言背景和文化水平密切相关。例如，“里弄”、“路牙子”等词可能随着时代变迁不为年轻人熟知，造成不同年龄群体的语义可接受性判断存在一定差异。本文更关注语言层面的共性问题，此不作讨论。

二、影响空间语义可接受性的认知因素

①参照物形态特征

参照物和方位词应在形态特征上相契合，否则无法形成参照域。例如，“空气、声音”因天然的包裹性被概念化为容器，能与指示内部的方位词组合，如6a的“空气中”；同时，又因缺乏可见边界和固定形体，无法自然支持“上下前后左右边旁外”等方位的表达，如6b的“空气边”就不可理解；

(6) a. 丝线的颜色也在空气中发生着反应，变得异常鲜艳。

b. *丝线的颜色也在空气边发生着反应，变得异常鲜艳。

除非在语境中被赋予新的特征，如例7的“一团热空气”在认知上获得了可区分的形状和边界，从而使“空气下面”这一表达可以被接受。

(7) 一团热空气在外力作用跑到了冷空气下面，两者会分层呆在不同高度。

二、影响空间语义可接受性的认知因素

①参照物形态特征

还有“天空、地面”这类被概念化为平面的实体，有可视范围但无朝向区分，能与方位词表达上下、里外、边缘等概念，如8a的“天空中”；但不能表达前后和左右，如“天空后”这个区域不存在。即便在语境中，也很难以“天空”为参照物构建前后左右等参照域，因为“天空”在认知上有无限的尺度，不能被重新概念化为一个有朝向的对象。

- (8) a. 有一颗星星落下去了，在天空中划出了一道细长的红光。
b. *有一颗星星落下去了，在天空后划出了一道细长的红光。

二、影响空间语义可接受性的认知因素

①参照物形态特征

此外，一些参照物虽然具有固定边界，但其仅支持单一的空间划分，因而也有不能表达的方位。例如，“纬度”只体现南北方向，却不能承载东西方向，9b的“纬度最东”无法成立；“脚掌”按照身体朝向区分前后部分，而没有上下部分，10b的“下脚掌”也不可接受。

(9) a. 四川盆地是中国著名红层盆地，是中国各大盆地中**纬度最南**、海拔最低的盆地。

b. *四川盆地是中国著名红层盆地，是中国各大盆地中**纬度最东**、海拔最低的盆地。

(10) a. 身体呈运动姿站立，双脚间距略比肩宽，手臂呈前后摆臂状，重心位于**前脚掌**。

b. *身体呈运动姿站立，双脚间距略比肩宽，手臂呈前后摆臂状，重心位于**下脚掌**。

二、影响空间语义可接受性的认知因素

②参照物数量特征

“间/中间/之间/之中”等方位词要求参照物必须具备复数性，即由两个或两个以上的实体构成。例如，“圆柱间”意味着至少存在两根圆柱，只有当“圆柱”被理解为群体性名词时，这一参照域才可理解。如果“圆柱”加上数量词“一根”，则因参照物单一而与“间”要求的复数性冲突，导致语义不可接受。

- (11) a. 我默默站在东边一根圆柱旁，没有看周围站着的人群。
b. *我默默站在东边一根圆柱间，没有看周围站着的人群。

二、影响空间语义可接受性的认知因素

(2) 空间介词+参照域

空间介词与参照域组合进一步定义场域的性质，是一个固定区域（位置义、距离义介词）还是一条位移路径（起点义、方向义、终点义、经由义介词）。

经由义介词和距离义介词对参照域预设特定的空间维度，若二者不匹配，则会导致语义上的不可接受性。

二、影响空间语义可接受性的认知因素

(2) 空间介词+参照域

经由义介词“沿/沿着/顺着”后接实体名词时，实体的线性维度特征被显化。如例12的“地毯”应理解为一个线性延展的平面；例13的“沿着桌子”标记了桌子的边缘。

(12) 周恩来总理在热烈的欢呼声中**沿着红色地毯**走向机场大厅。

(13) 桌上放了一本银色外皮的法术书，银粉**沿着桌子**在地上画了一个**圈圈**。

后接方位短语时，方位词应同样突显线性维度特征，否则不可接受。

(14) 两个人坐了一刻，又站起来**沿着湖边**走了几步。

(15) *文教干事正皱着眉头在看着一只矮下身子打算**沿一个孩子的凳上跳**的母鸡。

(16) *母企鹅只得张开嘴，让小企鹅把嘴伸到它嘴里，吃它**沿胃里**呕出来的食物。

二、影响空间语义可接受性的认知因素

(2) 空间介词与参照域

距离义介词“离/距”后接实体名词时，实体上距离测量点最近的位置被抽象为参照点。如例17并不关心“桌面、桌子、鼻尖”的大小、形状等属性，而是统一被抽象为计算距离的最近点。

(17) 眼睛离桌面一尺，身子离桌子一拳远，手指离笔尖一寸。

后接方位短语时，方位词应突显参照域的边界，从而在边界上抽象出最近点。

(18) 原来滨湖的多伦巴加社区，现在距离湖边已20多公里。

(19) *李天来离钻研新技术的路上越走越远。

二、影响空间语义可接受性的认知因素

(3) 位移路径

位移路径由起点、终点、经过点、方向、轨迹等部分组成。只有协调一致，才能组成路径。

①一致性

(20) a. 他跳入洪水里想把放映机抢捞上来。

b. *他跳入洪水外想把放映机抢捞上来。

(21) a. 他在幽静的小路上散步，听到断断续续的钢琴声从茅屋里传出来。

b. *他在幽静的小路上散步，听到断断续续的钢琴声向茅屋里传出来。

(22) a. 那雾像被谁牵系着，从县城上空掠过，速度快得惊人。

b. *那雾像被谁牵系着，向县城上空掠过，速度快得惊人。

二、影响空间语义可接受性的认知因素

②位移性

目标物在路径中必然经历从一个地方出入的变化，否则违背位移性要求。例如，23b的“在办公室里出来”就是一个不可接受的表达。

- (23) a. 保罗从办公室里出来的时候，看见一个男孩在他闪亮的新车旁走来走去。
b. *保罗在办公室里出来的时候，看见一个男孩在他闪亮的新车旁走来走去。

③唯一性

位移路径单向展开（返程视为两条路径），因而起点、经过点和终点必须在空间上不重合。

- (24) a. 臧晓梅驾驶小型客车由南向北行驶。
b. *臧晓梅驾驶小型客车由北向北行驶。

二、影响空间语义可接受性的认知因素

(4) 目标物与场域

目标物能否处于场域中，主要与以下几个方面有关。

①数量特征

目标物应有足够的数量去占据场域。

(25) *后排空间小，我坐左边，辛荑坐在右边，小红坐在我们两边。

②容纳能力

场域具备可进入性，同时目标物的体积也不能超过场域所能承受的限度。

(26) *大约是在课上到四十分钟时，一只母鸡在一个男孩的腿内停住了。

二、影响空间语义可接受性的认知因素

③客观约束强度

现实世界中不可改变的空间关系。

(27) **直肠在大肠的最上端，直接与肛部连接。*

(28) **宋钢茫然地转向窗外，看着马路对面的广告牌，看着广告牌上面的楼房，看着楼房下面的天空。*

④场域功能

场域基于参照物功用被赋予的用途。

(29) **她对着自己在镜子旁的样子出了半天的神。*

二、影响空间语义可接受性的认知因素

(5) 事件与固定区域

事件既占据时间，也占据空间。动词蕴含的事件语义与场域之间同样存在匹配约束。

①事件的空间适应性

储泽祥（1998）根据范围大小将空间区域分为三级：“难以承载或容纳人体、可承载或容纳多个人体、可容纳很多人体”。动词的空间适应能力越强，所能搭配的空间区域范围就越大。

(30) a. 他坐在凳子上。

b. 他坐在门口。

c. *他坐在日本。（崔希亮，1996）

二、影响空间语义可接受性的认知因素

②事件的位置性需求

事件可能对空间论元提出位置性要求，如果场域与该特征构成对立，则句子无法成立。以“跟”为例，意思是“在后面紧接着向同一方向移动”，充当空间论元的场域不能与“在后面”冲突。

- (31) a. 宋钢像是被打败了，十分窝囊地**跟在**（李光头）**后面**。
b. 宋钢像是被打败了，十分窝囊地**跟在**（李光头）**旁边**。
c. *宋钢像是被打败了，十分窝囊地**跟在**（李光头）**前面**。

二、影响空间语义可接受性的认知因素

③事件的功能性需求

事件可能对空间论元提出功能性要求，如“坐、放、载、摆”等动词包含“由某一表面支撑目标物”的语义，要求空间论元指示的场域提供一个承载面。

功能性需求有一定的灵活性，有时从常识或语境中能得到补偿。以“摆”为例，储泽祥（1998）谈到“墙上摆着”不能说，因为“摆”要求不能是侧平面。但从语料库中还是能找到一些用例：

(32) 墙上摆着的这些酒瓶每个都长得不一样，都是各种大神画的插画！



二、影响空间语义可接受性的认知因素

(6) 事件与位移路径

①事件的方向性

一些事件具有内在的方向性，如“举”天然向上， “踩”向下， “退”向后。当外在的方向表达不能与事件的内在方向兼容时，就会造成语义冲突。

- (33) a. 宋钢把剩下的饭菜全部塞进自己嘴里，他的嘴巴像个皮球一样**鼓起来了**。
b. 宋钢把剩下的饭菜全部塞进自己嘴里，他的嘴巴像个皮球一样**鼓出来了**。
c. *宋钢把剩下的饭菜全部塞进自己嘴里，他的嘴巴像个皮球一样**鼓下去了**。

二、影响空间语义可接受性的认知因素

“过来、过去”等趋向动词还要求事件具有意志性。如“咽”、“呼（呼吸）”等依赖生理器官被动完成的动作，以及“掉（掉落）”、“沉”等受重力作用而缺乏其他施力作用的动作，均无法与“过来、过去”搭配。

(34) a. 她将馒头一片一片放进嘴里，慢慢地咀嚼，慢慢地咽下去。

b. *她将馒头一片一片放进嘴里，慢慢地咀嚼，慢慢地咽过去。

(35) a. 胡中元买来一大卷编织得很密的网，铺在林间空处的“腰”部。这样，即便鸟蛋和雏鸟掉下来，也不会摔坏。

b. *胡中元买来一大卷编织得很密的网，铺在林间空处的“腰”部。这样，即便鸟蛋和雏鸟掉过来，也不会摔坏。

二、影响空间语义可接受性的认知因素

②事件与其他路径元素

(36) a. 当地传统工艺铜锣不仅在泰国销售，还出口国外。

b. *当地传统工艺铜锣不仅在泰国销售，还出口国内。

(37) a. 李光头和宋钢摇摇头仍然坐在门槛上，于是他就抬脚从他们上面跨了出去。

b. *李光头和宋钢摇摇头仍然坐在门槛上，于是他就抬脚从他们下面跨了出去。

(38) a. 文教干事正皱着眉头在看着一只矮下身子打算从一个孩子的凳上离开的母鸡。

b. *文教干事正皱着眉头在看着一只矮下身子打算从一个孩子的凳上跳的母鸡。

二、影响空间语义可接受性的认知因素

(7) 观察者与位移路径

趋向动词“来”、“去”的使用取决于观察者的位置。

- (39) a. 三十九岁的扎西达吉从汽车站前朝我们跑来。
b. *三十九岁的扎西达吉从汽车站前朝我们跑去。

- (40) a. 我站在候机大厅的落地窗前，目光追随着那架正在起飞的飞机。只见它缓缓地向蓝天飞去，最终消散在远方的云层中。
b. *我站在候机大厅的落地窗前，目光追随着那架正在起飞的飞机。只见它缓缓地向蓝天飞来，最终消散在远方的云层中。

三、面向空间信息正误判断任务的语义标签设计

第一层 空间词语蕴含的空间关系类型

第二层 空间语义异常类型

第三层 空间语义理解主题

三、面向空间信息正误判断任务的语义标签设计

第一层

基于
335个
空间词语总结的
空间关系标签

大类14个	中类12个	小类78个
方位标记	-	位置、起点、终点、方向、朝向、轨迹、经过点、距离
形态标记	-	凹、凸、横、竖、直、弯、斜、环
高度方位	高\低	高处、顶部、高于 \ 低处、底部、低于
范围方位	-	里边、外边、边缘、端部、周边、表面、中间、中心、周遍
地极方位	单向\复合	东边、西边、南边、北边 \ 东南、东北、西南、西北
身体方位	前\后\左\右\朝向\复合	前边、首部 \ 后边、尾部 \ 左边 \ 右边 \ 正面、反面、侧面、正面朝上、正面朝下、... \ 左前、左右、...
布局方位	-	垂直、平行、交叉、重叠
距离方位	-	近距离、远距离
高度变化	-	从低到高、从高到低、起伏
范围变化	-	从里到外、从外到里
距离变化	-	从远到近、从近到远
位置变化	方向\轨迹	从前动到后、从后动到前 \ 从终点到起点、从直行到转弯
朝向变化	-	扭转

三、面向空间信息正误判断任务的语义标签设计

第二层 空间语义异常标签

大类3个	中类16个	小类16个	微类18个
选择限制性	场域	参照物-方位词、空间介词-方位短语	形态、结构、数量、维度
	位移路径	方向-终点、方向-方向、起点-方向、起点-终点、经过点-终点、位置-方向	范围、高度、维度
	固定区域	目标物-位置、位置-距离	数量、距离
	事件	事件-位置、事件-方向、事件-终点、事件-起点、事件-经过点	范围、朝向、高度、前后、距离、形态、尺度、支撑性、容纳性、可视性、反复性、意志性、离散性、抽象性
	观察者	视角-方向	
语义合理性	容纳能力、区域功能、专业知识、常规空间关系		
语篇连贯性	位置验证、路径验证、形态验证、数量验证、流程构建、类比推理		

三、面向空间信息正误判断任务的语义标签设计

第三层 贯通不同语体的空间语义理解主题标签

以“路径走向”主题为例：

(41) 一条**南北向**的水泥板路两旁，各有一排杨树巍然挺立。在路的尽头，在**最东侧**的杨树下，竖立着一块碑，刻着“赵各庄抗日战争遗址”。

(42) 大西北的天然气管道一路向东，**西气东输**一线全长4142公里，西气东输二线全长8704公里，**横贯中国南北两端**，途中穿越戈壁、高山、草原、水网，施工难度为同类项目世界之最。

(43) 2019年4月16日8时54分许，被告人赵振兴驾驶牌号为鄂FAXXXX的小型普通客车**沿浦星公路**公交专用道**由南向北**行驶至浦放路南约300米处，适逢张某某驾驶电动自行车后座搭载被害人胡某某**沿此处人行横道线由北向南横穿浦星公路**，两车相撞。

四、大语言模型的评测结果分析

SpaGE2025 “空间信息正误判断任务” 共有3500道测试题。1269道“异常”标签，平均文本长度133词；2231道“正常”标签，平均文本长度97词。

模型	A R1-7b		B Qwen2.5-7b		C Qwen3-4b		D Qwen3-4b		E 多模型投票		F 合成+微调+投票	
	正常	异常	正常	异常	正常	异常	正常	异常	正常	异常	正常	异常
ACC.	91	36	93	14	90	33	96	19	85	41	71	58

模型	Deepseek-R1	
标签	正常	异常
ACC.	89	71

1. 模型在两个类别的表现存在显著不对称：模型对“正常”类的判断准确率很高，但对“异常”类的准确率很低。
2. 数据微调能有效提升“异常”类得分，但削弱了“正常”类得分。
3. 模型识别异常的能力可能随模型参数量增大而增强。

四、大语言模型的评测结果分析

语义异常题目分析

标签	题量	前六平均ACC.	R1
高度方位	217	42.63	74.65
身体方位	291	32.93	70.10
范围方位	323	32.56	70.06
地极方位	92	41.85	64.13
布局方位	7	30.95	57.14
距离方位	24	30.56	54.17

标签	题量	前六平均ACC.	R1
标记类	36	36.63	88.89
方位类	954	35.80	70.05
变化类	298	24.72	69.13

标签	题量	前六平均ACC.	R1
高度变化	92	28.44	77.17
范围变化	82	21.14	71.95
距离变化	65	23.08	67.69
位置变化	42	24.21	61.90
朝向变化	17	29.41	35.29

四、大语言模型的评测结果分析

语义异常题目分析：与地极方位和朝向变化相关的主题

标签	题量	前六平均ACC.	R1
碰撞条件	26	35.26	57.69
路径走向	16	23.96	56.25
方向换算	36	36.11	47.22
调转方向	15	34.44	33.33

(44) *陶述发酒后驾驶小型普通客车沿上海市嘉定区新成路由北向南行驶至新成路、嘉罗公路路口，见前方有民警设卡检查，遂调头后沿新成路向南行驶。

(45) *宋钢看着林红跨上自行车，叮嘱她骑车要小心，林红点点头往西骑车而去，宋钢扭头往西走向了码头。

(46) *图书馆闭馆音乐响起时，林晓把借阅的文献塞进双肩包。她沿着旋转楼梯往下走，忽然听见前方传来管理员的声音：“同学你过来，你的校园卡掉在检索机旁边了！”她连忙转身折回去。

四、大语言模型的评测结果分析

语义异常题目分析

标签	题量	前六平均ACC.	R1
高	110	42.12	74.55
低	112	43.30	75.00

标签	题量	前六平均ACC.	R1
高处	94	41.67	74.47
顶部	13	48.72	69.23
低处	84	44.84	76.19
底部	24	37.50	66.67

标签	题量	前六平均ACC.	R1
前	103	30.10	70.87
后	101	32.51	76.24
左	29	42.53	62.07
右	32	31.77	56.25

标签	题量	前六平均ACC.	R1
东	27	33.95	59.26
南	24	54.86	70.83
西	32	37.50	56.25
北	24	37.50	66.67

四、大语言模型的评测结果分析

语义异常题目分析

标签	题量	前六平均ACC.	R1
选择限制性	259	34.68	78.76
语义合理性	251	39.64	73.31
语篇连贯性	759	30.70	67.19

标签	题量	前六平均ACC.	R1
方向-终点	30	41.11	96.67
事件-方向	46	36.23	86.96
参照物-方位	20	28.33	80.00
目标物-区域	13	39.74	76.92
事件-位置	96	32.29	72.92
起点-方向	16	37.50	62.50
视角-方向	10	15.00	50.00

四、大语言模型的评测结果分析

语义异常题目分析

标签	题量	前六平均ACC.	R1
选择限制性	259	34.68	78.76
语义合理性	251	39.64	73.31
语篇连贯性	759	30.70	67.19

标签	题量	前六平均ACC.	R1
常规关系	145	0.395402	0.689655
区域功能	40	0.283333	0.75
容纳能力	38	0.451754	0.815789
专业知识	28	0.488095	0.821429