

人工智能 汉语框架语义技术标注规范
第 2 部分：句子语义标注

2025 - 05 - 07 发布

2025 - 08 - 07 实施

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 标注内容 1

5 标注格式 2

6 标注流程 2

7 标注方法 3

附录 A（资料性） 框架信息示例 4

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由山西省工业和信息化厅提出、组织实施和监督检查。

山西省市场监督管理局对本文件的组织实施情况进行监督检查。

本文件由山西省物联网和人工智能标准化技术委员会（SXS/TC22）归口。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件起草单位：山西大学、太原科技大学、山西同方知网数字出版技术有限公司、山西六新产业技术研究院有限公司、山西省数字政府服务中心、太原如是达科技有限公司。

本文件主要起草人：李茹、梁吉业、赵红燕、郭少茹、王长征、薛义飞、柴清华、苏雪峰、王智强、谭红叶、张虎、马志红、吕国英、高俊杰、闫智超、李俊材、王俞智、胡志伟、刘全明、王灵梅。

人工智能 汉语框架语义技术标注规范

第 2 部分：句子语义标注

1 范围

本文件规定了汉语框架句子语义标注的术语和定义、标注内容、标注格式、标注要求和标注方法等要求。

本文件适用于汉语框架句子语义标注的数据标注方开展训练数据标注活动。汉语框架句子语义数据需求方对于数据标注进行检查、验收或第三方机构对数据标注进行安全性评估的活动可参照执行。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件,仅该日期对应的版本适用于本文件;不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

DB14/T 3442—2025 人工智能 汉语框架语义技术标注规范 第 1 部分：框架构建

3 术语和定义

DB14/T 3442—2025 界定的以及下列术语和定义适用于本文件。

3.1

句子语义标注

依据框架语义对句子中词或短语的语义角色进行标记的过程。

3.2

短语

语义和语法上都能搭配的两个或更多的词的组合。

3.3

连谓结构

谓词或谓词结构连用的格式。

4 标注内容

句子语义标注的内容应包括目标词所激起的框架,与目标词语义相关的句子成分及其对应的框架元

素（语义角色）。标注示意图见图 1。

注：对于目标词相关的语言成分，其语义角色为目标词所激起框架下的框架元素。

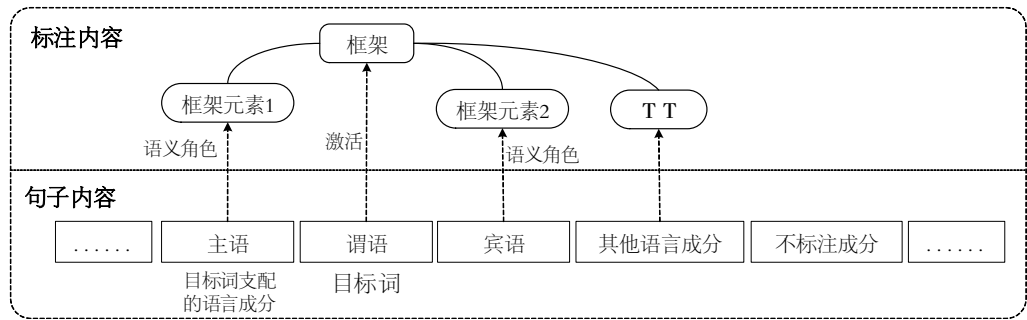


图 1 句子语义标注内容示意图

5 标注格式

5.1 标注句子时分别标注目标词以及目标词相关的句子成分，标注格式如下：

——目标词：<tgt=框架名称 目标词>；

——句子成分：<框架元素缩写 标注成分>。

示例：<owner 每个人> 都 <tgt=拥有 怀着> <possio 自己的梦想>，这些梦想都是他们努力前行的动力。

注：【拥有】框架信息见附录A中表A.1～表A.3。

5.2 标注格式应满足：

- a) 目标词激起的框架用“tgt=”标记；
- b) 每个被标注成分和对应标注的元素使用“<>”（尖括弧）括起；
- c) 框架元素用其缩写表示；
- d) 框架元素缩写和被标注成分之间空一格；
- e) 框架和目标词之间空一格；
- f) “<”与左边原文文字之间空一格；
- g) “>”与左右原文文字之间各空一格。

6 标注流程

6.1 汉语框架下的句子语义标注流程应按照图 2 所示实行。开始标注时，对于给定的句子，应先确定待标注的目标词，随后根据上下文语义确定其激起的框架，最后根据框架标注句子中的语义角色并结束标注。当句子中出现多个目标词时，应分别对每一个目标词进行标注。

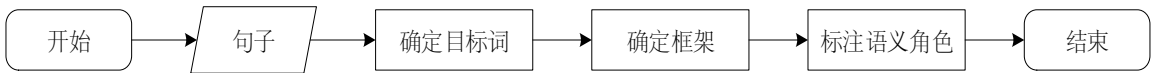


图 2 句子语义标注流程图

6.2 流程中以下环节应符合的具体要求为：

- 确定句子中的目标词。应根据已有的框架，确定句子中所有能够激起这些框架的目标词；
- 确定目标词激起的框架。应根据目标词的图式化场景，选定与其一致的框架；

——标注语义角色。应根据目标词所激起的框架，为句子中目标词相关的语言成分标注其对应的语义角色。

7 标注方法

7.1 标注范围

与目标词语义相关的所有句子成分都应标注。

7.2 特殊标注情况

在连谓结构中，标注不同谓词时应考虑谓词之间的语义关系并给定特定的语义角色。具体包括但不限于以下情况：

- 在连谓结构中，若该结构中的动词有先后关系，且后者是前者的目的，当前者做目标词时，后者应标注为“目的”；
- 在连谓结构中，若该结构中存在两个动词互为伴随动作，当其中一个动词做目标词时，另一个应标注为“并行事件”；
- 在连谓结构中，若存在动词 v1 为“领、带、带领、陪同、陪”等表示携带、陪同类动词，当其后存在动词 v2 且 v2 做目标词时，动词 v1 及其宾语应标注为“伴随者”；
- 在包含两个谓词的连谓结构中，当第一个动词是“来/去”，且“来/去”做目标词时，第二个动词应标注为“目的”；
- 在包含两个谓词的连谓结构中，若第一个动词为“有”且宾语为表示抽象事物的名词，如：办法、权利等，当第二个动词做目标词时，前者及其宾语应标注为“根据”。

7.3 无需标注情况

如遇词汇无实际意义或框架无法表示的情况不应标注。具体包括但不限于以下情况：

- a) 连谓结构中，出现超过两个谓语时；
- b) 当以“去”为目标词，且其在句子中无实际意义时；

示例：值得注意的是，知识分子只会批评，但又不开机看，而电视台一向以收视率为依据，要他们主动 <tgt= 去 > 改变这种状况几乎是不可能的。

- c) 连词、关联副词、语气词；
- d) 单纯表示否定的副词；
- e) 在包含两个谓词的连谓结构中，若第一个谓语动词为“领、带、带领、陪同、陪”等表示携带、陪同类动词，其做目标词时，其他的谓语结构不标注；
- f) 在包含两个谓词的连谓结构中，若第一个谓语动词为“有”且宾语为表示抽象事物的名词，如：办法、权利等，当“有”做目标词时，激起的框架应为【拥有】框架，另一个动词不标注；
- g) 在包含两个谓词的连谓结构中，若第一个谓语动词为“来/去”，当另一谓词做目标词时，“来/去”不标注；
- h) 偏古文的句子；
- i) 当目标词无法激起合适的框架时。

附录 A
(资料性)
框架信息示例

【拥有】框架信息示例见表 A.1 ~ 表 A.3。

表 A.1 【拥有】框架——框架信息

框架中文名称	框架英文名称	场景定义
拥有	Possession	所有者有（或没有）某种东西（拥有物），拥有物有可能是财产、资金、产权、地产、发明专利等。

表 A.2 【拥有】框架——框架元素信息

中文名称	英文名称	缩写	类型	框架元素定义
所有者	Owner	owner	核心	拥有某物（所有物）的实体。
拥有物	Possession	possio	核心	所有者拥有的东西。
形容	Depictive	depic	非核心	对所有者或拥有物状态的描述。
方式	Manner	manr	非核心	描述事件所处的状态。
时量	Duration	dura	非核心	一个可持续状况从开始到结束的时间长度。
解释	Explanation	expla	非核心	所有者拥有拥有物的原因。

表 A.3 【拥有】框架——词元信息

词元	词性	例句
有	动词	他 有 五辆汽车。
具有	动词	它 具有 粗生易长，适应性强的特性。