

CCL25-Eval任务3总结报告: 第五届中文抽象语义表示解析评测

许智星^{1,2}, 张艺璇^{1,2}, *李斌^{1,2}, 徐静^{1,2}, 曲维光^{2,3}, 周俊生³

1. 南京师范大学文学院
 2. 南京师范大学语言大数据与计算人文研究中心
 3. 南京师范大学计算机与电子信息学院
- libin.njnu@gmail.com

摘要

本文为第五届中文抽象语义表示解析评测 (CAMRP 2025) 的总结报告。CAMRP 2025包含两个子任务: 中文抽象语义表示 (CAMR) 句子级解析任务, 和CAMR篇章共指解析任务。评测任务共有96支队伍报名, 4支队伍提交结果, 最终总计26份有效成绩。哈尔滨工业大学(深圳)团队在开放测试下, 取得了84.72%的F值, 为CAMRP评测系列五年来的历史最好成绩。该团队在篇章共指消解任务中同样获得了最高61.15%的好成绩, 相比baseline有较大提升。参赛队伍的实验结果表明, 尽管基于监督微调 and 图聚合的策略在句子级解析任务中展现出了较好的性能, 但大模型对于细粒度的篇章共指关系识别仍然存在挑战。如何有效利用CAMR结构化信息来提升大模型篇章共指解析的性能, 仍是未来研究的重要方向。

关键词: 抽象语义表示; 语义解析; 篇章共指; 中文信息处理

Overview of CCL25-Eval Task 3: The Fifth Chinese Abstract Meaning Representation Parsing Evaluation

Zhixing Xu^{1,2}, Yixuan Zhang^{1,2}, *Bin Li^{1,2},
Jing Xu^{1,2}, Weiguang Qu^{2,3}, and Junsheng Zhou³

1. School of Chinese Language and Literature, Nanjing Normal University, China
 2. Center for Language Big Data and Computational Humanities,
Nanjing Normal University, China
 3. School of Computer and Electronic Information, Nanjing Normal University, China
- libin.njnu@gmail.com

Abstract

This paper summarizes the 5th Chinese Abstract Meaning Representation Parsing Evaluation (CAMRP 2025), which consists of two sub-tasks: the Chinese Abstract Meaning Representation (CAMR) sentence-level parsing task and the CAMR discourse-level coreference resolution task. A total of 96 teams registered for the evaluation, with 4 teams submitting results, resulting in 26 valid submissions. The team from Harbin Institute of Technology (Shenzhen) achieved an F1 score of 84.72% in the open modality, setting the highest score in the history of the CAMRP evaluation series. The team also achieved the highest score of 61.15% in the discourse coreference resolution task, showing a significant improvement over the baseline. It is observed that the strategy based on supervised fine-tuning and graph aggregation performed well in sentence-level parsing tasks. However, large models still face challenges in identifying fine-grained discourse coreference relations. Effectively leveraging the structured information of CAMR to enhance the performance of large language models in discourse-level coreference parsing remains an important direction for future research.

Keywords: Abstract Meaning Representation, Semantic parsing, Discourse coreference, Chinese information processing

1 引言

语义分析是目前自然语言处理的重点和难点，如何利用高质量的形式化表达，来解决自然语言中的语义模糊和歧义，一直以来都是国内外研究热点。尽管诸如DeepSeek和ChatGPT等大语言模型的出现，一定程度上以端到端的方式直接表征语义，在NLP多个下游任务中取得了显著进展。然而，目前大模型仍存在生成语言流畅而不精准、可解释性较差以及在细分领域的性能表现不足等问题。而前些年兴起的抽象语义表示（Abstract Meaning Representation, AMR）(Banarescu et al., 2013)通过图结构的方式来刻画一个句子的语义结构，则具有较强的语义表示能力和更好的可解释性。同时，中文AMR（Chinese Abstract Meaning Representation, CAMR）(Li et al., 2019)在AMR基础上，针对汉语语言特点进行了创新与改进，构建了较大规模的CAMR语料库，并在国内外举行了四届CAMR语义解析评测，取得了较好的成绩。因此，继续推动CAMR解析技术，可为优化大模型提供新的思路。

然而，当前的CAMR解析大多聚焦于句子级别的语义表示，忽略了篇章级别（Discourse-level）的语义关系和共指解析等复杂问题。篇章级语义解析不仅能够捕捉句与句之间的深层语义联系，还能提升文本理解的准确性和一致性。为了解决这一问题，推动CAMR从句子级解析向篇章级解析的扩展成为当前研究的一个重要方向。本文首先对近年来CAMR语义解析的研究进展进行了简要概述；其次，基于第五届CAMR解析评测任务的比赛结果，重点讨论CAMR在篇章共指解析中的应用和挑战。此外，还讨论并分析了现有的技术框架和评测方法，探讨未来研究中可能采用的新思路和方法，以期CAMR解析的进一步发展提供参考。

2 评测任务

为了更好地区分句子级与篇章级语义解析，本届CAMR解析技术评测分别设置了两个子任务，具体如下：

- 子任务A：该任务为CAMR句子解析任务，要求给定分词后的句子，解析系统输出对应的CAMR图结构，结果中需包含概念对齐和关系对齐信息。CAMR句子级解析一直是往届评测任务中的核心内容，本届继续保留该任务，以便更好地对比近年来CAMR解析技术的进展与提升。
- 子任务B：该任务为CAMR篇章共指消解任务，要求给定包含若干句子组成的篇章，解析系统输出该篇章中存在的所有共指链，结果中需包含共指关系、句子编号和共指词信息。具体而言，给定输入的篇章文本，参赛队的解析系统需要识别并输出该篇章中的所有共指链，并按照CAMR篇章共指标注的格式进行表示。

3 数据集

3.1 数据规模

中文抽象语义表示语料库（Chinese Abstract Meaning Representation Corpus）由南京师范大学与美国布兰迪斯大学于2015年联合构建，并于LDC（Linguistic Data Consortium，语言数据联盟）上以CAMR v2.0¹的形式发布。该语料库包含20078条中文句子，原始文本选自宾州中文树库（Chinese Treebank 8.0, CTB 8.0），并划分为训练集、验证集与测试集。CAMR v2.0已在CoNLL 2020及CAMRP 2022至CAMRP 2024 (Xu et al., 2023) (Xu et al., 2024)中多次用于评测，成为CAMR解析的重要基准资源。本届评测任务将继续使用该语料，以系统比较近两年来CAMR自动解析技术的进展。

与此同时，为进一步推动句子级语义表示向篇章层级扩展，本届评测任务新增500篇篇章级语料，用于篇章共指链的解析任务。新增语料取自CTB编号为chtb0001-chtb0659的子集，共计6,237句，涵盖经济、体育、生活等多种题材，具有较强的语言多样性与代表性。其中，300篇用作验证集（Dev B），200篇用作测试集（Test C），用于评估系统在篇章共指解析方面的能力。

©2025 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

¹<https://www ldc.upenn.edu/>

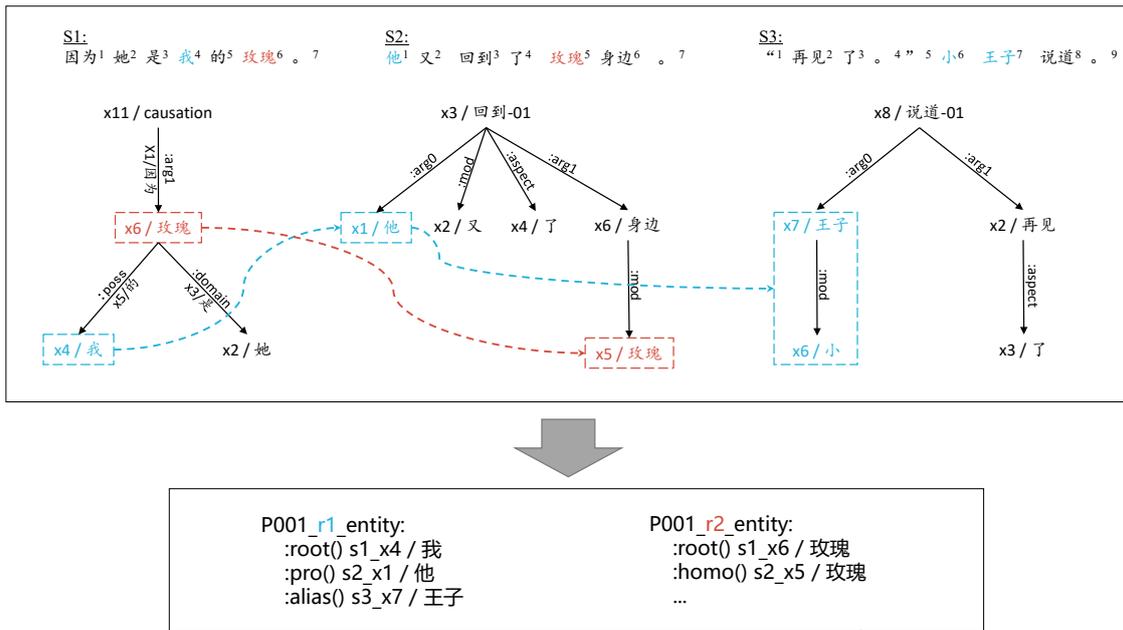


Figure 1: CAMR篇章共指标注样例图

Table 1展示了本次评测所用语料数据的详细信息。其中，训练集、验证集A和测试集A源自CAMR v2.0；测试集B为南京师范大学文学院自行标注的1999句人教版小学语文语料，用于评测解析系统的稳定性；验证集B为新增的300篇章共指标注，共计3581句，测试集C为新增的200篇章共指语料，共计2658句，用于评估系统在篇章共指解析方面的性能表现。

数据集	句子数	词例数
训练集	16,576	386,234
验证集A	1,789	41,822
验证集B (新增篇章)	3,581	96,247
测试集A	1,713	39,228
测试集B	1,999	36,940
测试集C (新增篇章)	2,658	71,724

Table 1: 评测语料详情

3.2 数据格式

子任务A中的数据格式沿用了往届的标准，包括CAMR文本表示、CAMR多元组表示以及依存句法分析结果。具体细节可参考CAMR任务系列的GitHub主页²。子任务B作为本届CAMR解析评测任务的创新，我们提供了全新的CAMR篇章共指标注语料(张艺璇et al., 2025)。Figure 1展示了CAMR篇章共指标注样例，上半部分为样例篇章中的三个句子(S1, S2和S3)以及对应的CAMR图示，虚线代表该篇章中存在的两条共指链，蓝色为“我-他-小王子”，红色为“玫瑰-玫瑰”；下半部分为该两条共指链对应的CAMR篇章共指标注文本样例。具体来说，CAMR篇章共指标注沿用了AMR的单根有向无环图结构，其中，“P001”为篇章编号，“r1”为共指链编号，“entity”为共指链类型，“:root”代表该词为先行词，“:homo”则表示该词与先行词之间的共指关系为“同形”，“s1”为共指词/先行词所在句子编号，“x4”为该词在原句中的词编号。

²<https://github.com/GoThereGit/Chinese-AMR>

4 评测流程

本届CAMRP 2025评测任务于2025年3月1日开放报名，并于4月10日上线天池平台³。评测所使用的训练集和验证集均由LDC官方提供，参赛队伍需自行签署评测语料许可证并向LDC发送邮件以获取评测语料。测试集A和测试集B先后于4月12日和4月25日分别发布于天池平台；5月18日天池平台关闭提交结果入口。5月20日至6月15日为参赛队提交技术报告及评测论文审稿环节，并于7月1日前完成Camera-Ready终稿提交，收录ACL/CCL会议论文集。CAMRP 2025评测任务将于2025年8月11日至14日在第二十四届中国计算语言学大会技术评测研讨会上举办，会议地点为山东济南，承办方为齐鲁工业大学（山东科学院）。

4.1 评测指标

子任务A的主要评测指标为Align-smatch (Xiao et al., 2022)，同时提供Smatch (Cai and Knight, 2013)与MRP (Abzianidze et al., 2020)指标作为参考，以详细对比近些年来CAMR解析的进展表现。该三种评测指标差异主要在于是否包含CAMR中的概念对齐和关系对齐信息(李斌et al., 2023)。子任务A中最终得分排名以Align-smatch下的F值为准。

子任务B的评测指标根据颗粒度不同，可分为Unlabel和Label两种模式：

- 在Unlabel模式下，我们仅对共指词进行匹配和检测，此时有：

$$TP = \sum_i (\text{coreferent}_i^{\text{parsed}}, \text{coreferent}_i^{\text{gold}}) \quad (1)$$

同理如FP和TN，最终返回 F_1 值为参考。

- 在Label模式下，我们需对共指链中的共指关系、句子编号和共指词进行匹配和检测，该字段构成一个共指三元组，此时有：

$$TP = \sum_i (\text{tuple}_i^{\text{parsed}}, \text{tuple}_i^{\text{gold}}) \quad (2)$$

同理返回 F_1 作为最终得分。

4.2 评测模态

本次评测任务包含开放测试（Open Modality）和封闭测试（Closed Modality），两种模态下的要求如Table 2所示：

- 若参赛队选择封闭测试，则必须使用指定的训练集、测试集和预训练语言模型，不可自行替换为别的资源。在封闭测试中，主办方提供训练集的依存分析结果，并推荐使用哈工大的HIT_Roberta预训练模型(Cui et al., 2021)。
- 若参赛队选择开放测试，预训练语言模型可自由选择（包含ChatGPT等），允许使用外部资源，如专名识别、依存句法分析结果等。开放测试中，参赛队使用的所有资源需要在最终提交的技术报告中给予详细说明。但无论哪种模态，均不可使用人工修正自动解析结果的方式。

5 评测结果

5.1 参赛信息

本届评测任务共计96支队伍报名，由于任务复杂难度较高，最终共有4支队伍提交结果42份。其中，1支队伍成绩无效，因而未纳入后续结果统计与分析。Table 3展示了参赛队伍的详细信息，排名按队伍名首字母而非代表最终名次先后，具体包含队伍名称、学校机构名称以及提交结果份数。参赛情况总体与上一届评测任务相似，所有队伍均选择开放测试。所有参赛队伍均在5月18日截止时间前完成提交，无逾期提交结果且无人工修正结果。

资源 \ 模态	模态	
	<i>Closed</i>	<i>Open</i>
算法	无限制	无限制
模型	HIT_Roberta	无限制
外部资源	依存分析结果	无限制
数据集	指定训练集和验证集	无限制
人工修正	禁止	禁止

Table 2: 两种模态下的要求

队伍名	机构来源	Test A		Test B		Test C	
		<i>closed</i>	<i>open</i>	<i>closed</i>	<i>open</i>	<i>closed</i>	<i>open</i>
CUMT	中国矿业大学 (北京)	0	1	0	1	0	1
HITSZ	哈尔滨工业大学 (深圳)	0	5	0	5	0	1
THU	清华大学	0	5	0	5	0	2
总计	26 (份)	0	11	0	11	0	4

Table 3: 参赛队信息及提交结果统计

5.2 结果统计与分析

最终提交有效成绩的团队分别为CUMT, HITSZ和THU, 所有队伍均只参加了开放测试。我们分别采用了三种评测指标, 统计了三支队伍在两个子任务上的表现, 其中, 由于CUMT和THU团队提交的数据格式存在部分兼容问题, 因此无法统计MRP指标下的得分。

Table 4和Table 5分别展示了参赛队伍在测试集A和测试集B上的表现, 即子任务A的整体得分。CUMT团队由于训练语料获取出现了意外, 在两个测试集上的表现受到了较大影响。THU团队以GPT4o为基座模型, 采用Few shot的微调策略完成CAMR句子级解析。该团队在prompting中额外加入了概念对齐和关系对齐信息, 保证了CAMR生成结果的有效性, 最终在测试集A上获得了50.88%的 F_1 值, 在测试集B上获得了45.58%的 F_1 值 (Align-smatch指标下)。HITSZ团队参考了往届CAMRP评测任务其他队伍的思路, 基座模型选取Llama、GLM和Qwen等多个大模型, 使用监督微调(Supervised Fine-tuning, SFT)的策略, 并以图聚合的方法完成了CAMR解析。该团队在后处理阶段构建了一个多智能体系统 (Multi-Agent System), 较好地解决了解析系统的幻觉问题, 最终在测试集A和测试集B上分别拿到了本届评测任务的最高分82.03%和76.80% F_1 值。同时, 该成绩也是自CoNLL 2020以来连续五届CAMRP评测的SOTA成绩, 打破了SUDA团队(Gu et al., 2023)自CAMRP 2023以来在测试集A上保持的记录, 以及GDUFU团队(Wu et al., 2024)在测试集B上的最佳记录。在MRP指标下, HITSZ团队的得分84.72%也显著高于CoNLL 2020上的最好成绩 (80.52%) (Samuel and Straka, 2020)。

Table 6展示了参赛队伍在测试集C上的表现, 分别包含Unlabel与Label两个模式下的得分。我们同时也对DeepSeek-V3进行了Few shot微调实验, 以作baseline方便与其他团队成绩对比。对于子任务B篇章共指解析, THU团队同样选用和子任务A一样的基准模型GPT4o, 采用Zero shot微调策略, 在prompting中加入了共指关系类型的学习, 最终在两个模式下分别得分为12.33%和8.43%, 相较于baseline有一定提升。HITSZ团队则再次拿到了测试集C上的最高分: 该团队首先对CAMR完成序列化处理, 结合篇章文本进行拼接, 完成监督微调训练, 最终利用构建的多智能体系统有效实现了错误检测与纠正, 在Unlabel模式下达到了61.15%的 F_1 值, Label模式下为18.20% F_1 值。

总体来说, 三支队伍均采用了针对大模型的微调训练策略, 充分发挥了预训练大模型在语言理解中的优势。实验结果表明, 监督微调相较于Few-shot或Zero-shot方法, 能显著提高模型的表现, 尤其在句子级别的解析任务中, 取得了较好的精度。然而, 在面对较为复杂的篇章共指解析任务时, 大模型依然面临一定挑战。尽管通过微调可以有效地提升模型在句子级语义解

³<https://tianchi.aliyun.com/competition/>

队伍 \ 指标	ALIGN-SMATCH			SMATCH			MRP		
	P	R	F ₁	P	R	F ₁	P	R	F ₁
CUMT	4.86	20.34	7.84	23.70	7.32	11.19	-	-	-
HITSZ	81.37	82.70	82.03	79.96	81.30	80.62	84.06	85.38	84.72
THU	46.77	55.77	50.88	41.67	32.61	36.59	-	-	-

Table 4: 开放测试中参赛队伍在测试集A上的成绩(%)

队伍 \ 指标	ALIGN-SMATCH			SMATCH			MRP		
	P	R	F ₁	P	R	F ₁	P	R	F ₁
CUMT	12.77	18.56	15.13	20.31	13.54	16.25	-	-	-
HITSZ	76.73	76.87	76.80	76.81	77.41	77.11	80.32	81.15	80.73
THU	42.68	48.91	45.58	50.01	40.57	44.79	-	-	-

Table 5: 开放测试中参赛队伍在测试集B上的成绩(%)

队伍 \ 指标	UNLABEL			LABEL		
	P	R	F ₁	P	R	F ₁
HITSZ	75.28	57.85	61.15	16.86	21.67	18.20
THU	22.23	9.14	12.33	8.01	10.23	8.43
DeepSeek-V3 _{Few-shot}	7.07	6.15	6.28	3.26	3.71	3.29

Table 6: 参赛队伍以及基线模型在测试集C（篇章共指）上的成绩(%)

析中的表现，但在捕捉篇章中句子之间的语义关联方面，模型仍存在局限性。尤其是在处理跨句共指链时，大模型未能完全解决长文本中的上下文依赖问题，且在细粒度的共指关系识别方面仍有明显短板，最终导致共指解析效果未能达到理想精度。

6 总结与展望

本届CAMR解析评测任务的结果表明，基于微调策略的大模型在句子级解析任务中取得了显著进展。尤其是HITSZ队伍，通过监督微调和多智能体系统，实现了最高84.72%的F₁值，成功打破了五届CAMRP评测任务的SOTA成绩，在多个评测指标下均有较大提升。然而，在篇章共指解析任务中，尽管大模型通过微调训练取得了一定进展，但仍面临跨句共指链识别以及细粒度共指关系识别的挑战，可以说篇章级解析是CAMR发展的重要方向。因此，为进一步提升CAMR篇章共指解析的准确性，未来的研究将关注优化长文本的上下文建模技术，结合多模态和多任务学习方法，针对当前大模型的局限性，融合CAMR丰富的结构化信息，进一步提升模型的可解释性以及复杂语义关系和跨句共指关系的建模能力。

参考文献

- L Abzianidze, Stephan Oepen, Omri Abend, Lasha Abzianidze, Johan Bos, Jan Hajič, Daniel Hershcovich, Bin Li, Tim O’Gorman, Nianwen Xue, et al. 2020. Mrp 2020: The second shared task on cross-framework and cross-lingual meaning representation parsing. *Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing*, pages 1–22.
- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse*, pages 178–186.
- Shu Cai and Kevin Knight. 2013. Smatch: an evaluation metric for semantic feature structures. In

Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 748–752.

- Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-training with whole word masking for chinese bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3504–3514.
- Marco Damonte, Shay B Cohen, and Giorgio Satta. 2017. An incremental parser for abstract meaning representation. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 536–546.
- Zhengxiao Du, Yujie Qian, Xiao Liu, Ming Ding, Jiezhong Qiu, Zhilin Yang, and Jie Tang. 2021. Glm: General language model pretraining with autoregressive blank infilling. *arXiv preprint arXiv:2103.10360*.
- Yanggan Gu, Shilin Zhou, and Zhenghua Li. 2023. CCL23-eval 任务2系统报告:基于图融合的回归和非回归中文AMR语义分析(system report for CCL23-eval task 2: Autoregressive and non-autoregressive Chinese AMR semantic parsing based on graph ensembling). In Maosong Sun, Bing Qin, Xipeng Qiu, Jing Jiang, and Xianpei Han, editors, *Proceedings of the 22nd Chinese National Conference on Computational Linguistics (Volume 3: Evaluations)*, pages 53–63, Harbin, China, August. Chinese Information Processing Society of China.
- Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.
- Bin Li, Yuan Wen, Weiguang Qu, Lijun Bu, and Nianwen Xue. 2016. Annotating the little prince with chinese amrs. In *Proceedings of the 10th Linguistic Annotation Workshop held in Conjunction with ACL 2016 (LAW-X 2016)*, pages 7–15.
- Bin Li, Yuan Wen, Li Song, Weiguang Qu, and Nianwen Xue. 2019. Building a chinese amr bank with concept and relation alignments. *Linguistic Issues in Language Technology*, 18.
- Stephan Oepen, Omri Abend, Lasha Abzianidze, Johan Bos, Jan Hajic, Daniel Hershcovich, Bin Li, Tim O’Gorman, Nianwen Xue, and Daniel Zeman. 2020. Proceedings of the conll 2020 shared task: Cross-framework meaning representation parsing. In *Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing*.
- Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744.
- Hiroaki Ozaki, Gaku Morio, Yuta Koreeda, Terufumi Morishita, and Toshinori Miyoshi. 2020. Hitachi at mrp 2020: Text-to-graph-notation transducer. In *Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing*, pages 40–52.
- David Samuel and Milan Straka. 2020. Úfal at mrp 2020: Permutation-invariant semantic parsing in perin. In *Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing*, pages 53–64.
- Zihao Wu, Yin Hua, Ziqian Gao, Jiajia Zhang, Yuelei Ji, and Kuntian Tang. 2024. 混合LoRA 专家的中文抽象语义表示解析框架. In *Proceedings of the 23rd Chinese National Conference on Computational Linguistics (Volume 3: Evaluations)*, pages 143–153, Taiyuan, China, July. Chinese Information Processing Society of China.
- Liming Xiao, Bin Li, Zhixing Xu, Kairui Huo, Minxuan Feng, Junsheng Zhou, and Weiguang Qu. 2022. Align-smatch: A novel evaluation method for chinese abstract meaning representation parsing based on alignment of concept and relation. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5938–5945.

Zhixing Xu, Yixuan Zhang, Bin Li, Junsheng Zhou, and Weiguang Qu. 2023. Overview of CCL23-eval task 2: The third chinese abstract meaning representation parsing evaluation. In *Proceedings of the 22nd Chinese National Conference on Computational Linguistics (Volume 3: Evaluations)*, pages 70–83, Harbin, China. Chinese Information Processing Society of China.

Zhixing Xu, Yixuan Zhang, Bin Li, Zhou Junsheng, and Weiguang Qu. 2024. Overview of CCL4-eval task 4: The Fourth Chinese Abstract Meaning Representation parsing evaluation. In Hongfei Lin, Hongye Tan, and Bin Li, editors, *Proceedings of the 23rd Chinese National Conference on Computational Linguistics (Volume 3: Evaluations)*, pages 160–171, Taiyuan, China, July. Chinese Information Processing Society of China.

张艺璇, 李斌, and 许智星. 2025. 从句子图到篇章图——基于抽象语义表示的篇章级共指标注体系研究. *外语学刊*, 1:19–28.

李斌, 许智星, 肖力铭, 周俊生, 曲维光, and 薛念文. 2023. 第二届中文抽象语义表示解析评测. *中文信息学部*, 37:33–43.