

# 陈俊杰

junjiechen.chris@outlook.com | Homepage | LinkedIn | Github | Google Scholar

## 工作经历

### Amazon Science Fellow

Amazon

开发及实现整书级多模态翻译系统  
研发翻译系统的评价体系

2025 年 8 月至今

东京, 日本

### 研究助理 (兼职)

国立情报学研究所 (NII)

开发日语大模型的评测脚本

维护包含 800+ 张 H100 GPU 的训练集群, 保障系统稳定运行

2024 年 6 月 - 2025 年 2 月

东京, 日本

### 研究实习生

乐天株式会社 (Rakuten)

改进基于大模型的隐私信息识别系统

基于 LSTM, 使基于大模型的 SAE 模型提升了约 25% 的识别精度

2024 年 11 月 - 2025 年 2 月

东京, 日本

### 教学助理

东京大学

指导了三组学生期末作业, 其中一组的成果发表于 JNLP-2023

2022 年 10 月 - 2023 年 2 月

东京, 日本

## 专业技能

研究方向: 通过对语言模型的采样, 进行语法/语义结构分析

模型框架: VAE、LLM、Diffusion、基于语法的生成模型

机器学习工具: PyTorch、Huggingface Accelerate、Lightning、JAX、NumPy

语言能力: 日语 (N1, 商务水平), 英语 (TOEFL iBT 102, C1)

## 项目经历

### 提出字符串语义信息度量, 增强无监督语法分析的精度

[GitHub 链接](#)

通过无监督语法解析精度近似无精标数据情况下细粒度语义解析的精度。

解决了基于语言模型相似度方法中, 无法准确预测字符串级别语义相似度的问题。

提出了一种基于转述操作, 从大型语言模型中估计字符串语义信息的新方法。

提出的语义信息度量大幅提高模型的语法解析精度, 在中、英、法、德四种语言上平均准确率提升达 8%。

研究成果发表于 ACL 2024 与 ICLR 2025 (Spotlight)。

### 基于 Metroplis-Hasting 方法, 并融合语法先验的依存句法分析模型

[GitHub 链接](#)

使用 Metroplis-Hasting 方法, 提升从语言模型中预测词间互信息精度。

提出将外部语法先验融入互信息计算的方法, 大幅提升语法解析精度。

语法依存结构整体预测准确率提升超过 5%, 语义相关依存结构的准确率提升超过 10%

研究成果发表于 NAACL 2024。

## 基于混合专家模型的语义依存分析模型

[GitHub 链接](#)

使用混合模型，对不同语法结构分别进行语义依存关系的建模。  
使用变分推断，成功将类似语义信息量的语法结构进行聚类。  
相比最新基线模型，在语义依存分析任务上有显著的性能提升。  
研究成果发表于 ACL 2022 (Oral)。

## 日文汉字生成模型

[Colab Notebook](#)

引入 Negative Prompt 技术，有效提升生成样式的清晰度与辨识度。

## 字符级语言建模系统

[GitHub 链接](#)

引入“distance2whitespace”损失函数，加快模型收敛速度。

## 奖项与资助

---

日本学术振兴会 (JSPS) 杰出 DC 研究员 (2024–2025): 月生活补助 3 万日元

AAAI 2025 国际会议差旅资助 (2025): 2000 美元

日本学术振兴会 (JSPS) DC2 奖学金 (2023–2025): 月生活补助 20 万日元, 年研究经费 90 万日元

东京大学 IST-RA 奖学金 (2022–2023): 月生活补助 6 万日元

## 精选论文

---

**Junjie Chen**, Xiangheng He, Yusuke Miyao, and Danushka Bollegala. Improving Unsupervised Constituency Parsing via Maximizing Semantic Information *In proceedings of ICLR-2025 (Spotlight, top 5.1%)*

Xiangheng He, **Junjie Chen**, Zixing Zhang, Björn W Schuller. ProsodyFM: Unsupervised Phrasing and Intonation Control for Intelligible Speech Synthesis. *In proceedings of the AACL Conference on Artificial Intelligence: AACL-2025 (Oral)*

**Junjie Chen**, Xiangheng He, Danushka Bollegala, and Yusuke Miyao. Constituents are Frequent Word Sequences among Sentences with Equivalent Predicate-Argument Structures: Unsupervised Constituency Parsing by Span Matching. *In findings of the Association for Computational Linguistics: ACL 2024*

**Junjie Chen**, Xiangheng He, and Yusuke Miyao. Language Model Based Unsupervised Dependency Parsing with Conditional Mutual Information and Grammatical Constraints. *In proceedings of the 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics*

**Junjie Chen**. Syntactic-Semantic Dependency Correlation in Semantic Role Labeling: a Shift in Semantic Label Distributions. *Journal of Natural Language Processing, 29(3)*

**Junjie Chen**, Xiangheng He, and Yusuke Miyao. Modeling Syntactic-Semantic Dependency Correlations in Semantic Role Labeling Using Mixture Models. *In Proceedings of 60th Annual Meeting of the Association for Computational Linguistics (Oral)*

Xiangheng He, **Junjie Chen** and Bjorn W. Schuller. Task Selection and Assignment for Multi-modal Multi-task Dialogue Act Classification with Non-stationary Multi-armed Bandits. *In proceedings of 2024 International Conference on Acoustics, Speech, and Signal Processing*

## 教育背景

---

东京大学

博士, 计算机科学, 导师: [Yusuke Miyao](#)

利物浦大学

访问博士生, 导师: [Danushka Bollegala](#)

东京大学

硕士, 计算机科学, 导师 [Yusuke Miyao](#)

东北大学

学士学位, 软件工程

东京, 日本

2021 年 10 月 - 2025 年 3 月

利物浦, 英国

2023 年 10 月 - 2024 年 3 月

东京大学

2019 年 10 月 - 2021 年 9 月

沈阳, 中国

2015 年 9 月 - 2019 年 6 月