

# 先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	自然言語処理学 (渡辺 太郎 (教授))		
学籍番号	2211340	提出日	令和 6年 1月 18日
学生氏名	SUNG JUNEHWAN		
論文題目	Exploring Metalinguistic Awareness in Pre-trained Language Models through the International Linguistics Olympiad Challenges		
要旨			
<p>In the field of Natural Language Processing (NLP), recent advancements have opened new horizons for human-machine interactions. However, the degree to which these models replicate “human-like” linguistic cognition still remains uncertain. This thesis seeks to unravel the answer by exploring metalinguistic awareness, the capacity to discern the underlying structure of language, within Pre-trained Language Models (PLMs), which are neural networks trained on extensive linguistic data, or corpora.</p> <p>Our study involves challenges from the International Linguistics Olympiad (IOL), a renowned international contest assessing linguistic proficiency. Among its various types of questions, we focus on “Rosetta Stone” problems, tasking participants with translating an unknown language solely relying on minimal provided information.</p> <p>Given the nature of these problems as translation tasks, we selected two PLM architectures -- mT5 and ByT5 -- acknowledged for their excellence in translation tasks. Our findings reveal a nuanced insight: while these models adeptly analyse implicit linguistic patterns, their metalinguistic awareness remains limited, especially in zero-shot learning.</p> <p>Furthermore, our exploration extends to evaluating these models without their pre-trained knowledge. The results underscore the crucial role of pre-training in enabling effective analysis and decoding of unknown languages.</p>			