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	2411097	提出日	令和 8年 1月 19日
学生氏名	北野　雄士		
論文題目	Analyzing Word Embedding Representations Across Layers of Pre-trained Multilingual Models Using Independent Component Analysis		

要旨

Multilingual models, trained on vast amounts of multilingual text data, have enabled sophisticated cross-lingual processing. This performance is fundamentally supported by embeddings, the technology that transforms linguistic units into high-dimensional vectors for computational processing. In these models, multiple languages are typically mapped into a single, shared embedding space. However, the internal mechanisms governing how word meanings and language-specific features are numerically represented within the embedding space remain largely opaque.

This research aims to elucidate the internal structure of a multilingual translation model using Independent Component Analysis (ICA), a method designed to decompose complex, mixed data into statistically independent, meaningful components. By applying this technique, the analysis revealed that while embeddings in the early layers are primarily clustered by surface-level features, this organization shifts toward language-independent meanings in the later layers.

This results visualize how multilingual models transform simple textual information into universal semantic concepts by abstracting data across layers. This study offers significant insights into the fundamental mechanisms of linguistic understanding in multilingual models.