

数学・数理科学と情報科学の連携・融合による情報活用基盤の創出と社会課題解決に向けた展開

2021年度採択研究代表者

2021年度 年次報告書

田中 久美子

東京大学 先端科学技術研究センター
教授

自然言語の非線形性の計算論モデル

§ 1. 研究成果の概要

本研究では、複雑系班、計算言語班、数理論理班の三つの班があり、各グループにおいて個別の研究プロジェクトが立ち上げられつつある。

初年度は、体制を整え、研究プロジェクトの立ち上げを行った。各グループで、人員計画が定まりつつある。グループ A,C で雇用する特任助教が決まり、各グループの RA がおおまかに定められた。また、各グループの中で研究を進める研究者を見直した。グループ間の研究体制作りも行った。体制の整備は、引き続き2年目に進めていく。

立ち上げた個別研究プロジェクトから、最初の成果が見え始めている。本研究は、スケールに応じて異なる目標を設定しており、マクロ、メゾ、ミクロで2つずつ目標を設定している。特に、構文構造と意味の関係に関する成果(メゾ 2)に関して、国際会議採択が決まった。また、単語のベクトルの新しい表現の成果(ミクロ 1)、文の意味の合成性に関する成果(メゾ 2)についても、論文を投稿中である。この他、論文を準備中のものが複数あり、計画書に記載した予定にしたがって順調に進められている。

2月24日には、初回のADミーティングを行い、プロジェクトの応用について、アドバイスをいただいた。アドバイザーの平田先生、穴井先生に、全体像の確認の後、計算言語班の取り組みを紹介し、応用の創出にむけた前向きなアドバイスをいただいた。

また、研究代表者は2022年の4月1日以降、ケンブリッジ大にて在外研究を行う計画であり、その準備を行った。

§ 2. 研究実施体制

(1) 研究代表者グループ（複雑系グループ）

- ① 研究代表者: 田中 久美子（東京大学先端科学技術研究センター 教授）
- ② 研究項目
 1. 系列の非定常性の数理モデル(マクロ 1)
 - 言語の非定常性と非線形構造の関係 (マクロ 1-1)
 - カオス論的解析に基づく言語の特質(マクロ 1-2)
 - 物理モデル、隠れ系列モデル、深層学習との関係(マクロ 1-3)
 2. 言語の開放性と埋め込み表現(マイクロ 1)
 - 単語の開放性と系列の非定常性の関係(マイクロ 1-1)
 - サブワード表現と言語の開放性(マイクロ 1-2)
 3. 言語の非線形性の統一的理解(全体 1)、計算言語学の基礎の整理(全体 2)

(2) 共同研究グループ (1) (計算言語グループ)

- ① 主たる共同研究者: 宮尾 祐介（東京大学大学院情報理工学系研究科 教授）
- ② 研究項目
 1. 文構造の非線形性の数理モデルと構文解析への応用 (メゾ 1)
 2. 単語構造から文構造までの統合的数理モデル (マイクロ 2)

(3) 共同研究グループ (2) (数理論理グループ)

- ① 主たる共同研究者: 峯島 宏次（慶應義塾大学文学部 准教授）
- ② 研究項目
 1. 文と論理式の相互変換可能性(メゾ2)
 2. 証明構造としての文列の非線形性と計算論モデル(マクロ2)

【代表的な原著論文情報】

- 1) Hitomi Yanaka and Koji Mineshima. “Assessing the Generalization Capacity of Pre-trained Language Models through Japanese Adversarial Natural Language Inference”, Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, pp.337-349, 2021. DOI: 10.18653/v1/2021.blackboxnlp-1.26
- 2) Hitomi Yanaka and Koji Mineshima. “Compositional Evaluation on Japanese Textual Entailment and Similarity” ,Transactions of the Association for Computational Linguistics, submitted
- 3) Junjie Chen, Xiangheng He, Yusuke Miyao.”Modeling Syntactic-Semantic Dependency Correlations in Semantic Role Labeling Using Mixture Models” ,Proceedings of ACL 2022,accepted
- 4) Tu-Anh Tran, Yusuke Miyao.”Development of a Multilingual CCG Treebank via Universal Dependencies Conversion”, Proceedings of LREC 2022, submitted

- 5) Xin Du and Kumiko Tanaka-Ishii. “Semantic field of words represented as potential functions”, Proceedings of ICML 2022, submitted