

渡航の成果の詳細

笹野 遼平

情報学研究科 知能システム学専攻/価値創造研究センター 准教授

1. 渡航中の研究成果

2017年10月1日～2018年3月31日の日程で、ケンブリッジ大学の理論・応用言語学部の Language Technology 研究所に滞在し、言語横断的な語彙知識の獲得の研究に取り組んだ。まず、言語横断的な語彙知識の獲得の基礎技術として1つの言語を対象に、(a)特定の意味クラスに属する単語埋め込みベクトルの意味ベクトル空間における分布に関する研究、および、言語横断的な語彙知識の構築を目指し、(b)英語を対象に人手で構築されたフレーム知識である FrameNet と日本語を対象に自動獲得された語彙知識を対応付ける研究に取り組んだ。以下に詳細を記す。

(a) 特定の意味クラスに属する単語埋め込みベクトルの分布の分析

言語間の意味クラスの対応付けを行うための基礎技術として、1つの言語において特定の意味クラスに属する単語埋め込みベクトルが、意味ベクトル空間においてどのように分布しているかの分析、および、その意味クラスに属しているかどうかの識別タスクに取り組んだ。具体的には、特定の意味クラスに属する語のベクトル集合が与えられた場合に、そこに含まれない語がその集合の要素となりうるかを識別するタスクに取り組み、与えられた集合(正例)のみに基づくモデルとして、与えられたベクトル集合の平均ベクトルとの距離に基づくモデル(MEAN)、多次元ガウス分布に基づくモデル(GM)、ガウス混合分布に基づくモデル(GMM)、k近傍ベクトルの平均ベクトルとの距離に基づくモデル(KNN)の4つのモデル、与えられた集合以外の語(負例)も考慮したモデルとしてSVMに基づく識別モデル(SVM)、多層パーセプトロンに基づく識別モデル(MLP)の6つのモデルを比較することで、対象の意味クラスに属するかどうかの識別には対象のクラスに属さない単語の情報が重要であることを明らかにした。本研究は論文にまとめ国際会議に投稿中である。

(b) FrameNet と日本語語彙知識の対応付け

言語横断的な語彙知識の構築を目指し、英語を対象に人手で整備されたフレーム知識である FrameNet と、日本語を対象に自動獲得された語彙知識の対応付けに取り組んだ。本研究では、日本語を対象に自動獲得された語彙知識として京大格フレームを利用し、MUSE (<https://github.com/facebookresearch/MUSE>)を用いて構築した言語横断的な単語ベクトルを手掛りとして、FrameNet に記載された各フレームを喚起する語の集合と日本語の動

詞の類似度, および, FrameNet の情報が付与されたコーパスから得られるフレーム要素を埋める語の集合と, 対象の動詞の格フレームの各格スロットの用例の類似度を考慮することで, FrameNet と京大格フレームの対応付けを行う。また, 評価は日本語版フレームネット

(<http://jfn.st.hc.keio.ac.jp/ja/home-ja/>)を用いて行う。具体的には, 日本語版フレームネットに含まれるアノテーション済テキストを, 京大格フレームに基づく

述語項構造解析器である KNP(<http://nlp.ist.i.kyoto-u.ac.jp/?KNP>)で解析し, その結果を上記で得られた対応付け知識に基づき FrameNet 形式の意味役割表現に変換することで, 日本語を対象とした意味役割解析を行い, 人手で付与された情報と比較することで対応付け知識の評価を行う。現在, 日本語版フレームネットからの評価用データの抽出・整備, MUSE を用いた言語横断的な単語ベクトルの獲得を終え, FrameNet と京大格フレームの対応付けモデルの構築を進めている段階であり, 今後も引き続き共同研究を継続する予定である。研究の概略を図 1 に示す。

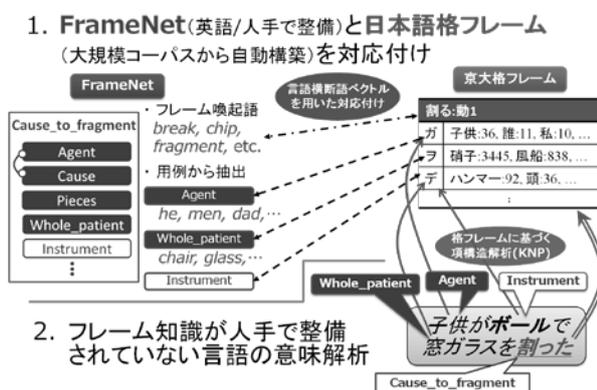


図 1. FrameNet と京大格フレームの対応付けと対応付け知識に基づく意味解析

2. その他の活動

- LTL Seminar で講演 (2017 年 11 月 2 日)

Language Technology Lab (LTL) で開催されたセミナーにおいて "Corpus-Based Analysis of the Canonical Word Order of Japanese Double Object Constructions" という題目で講演 (<http://talks.cam.ac.uk/talk/index/94753>)。

- LEXICAL プロジェクトのミーティング (2018 年 1 月 7,8,9 日)

LEXICAL (Lexical Acquisition across Languages) プロジェクトのミーティングにおいて「特定の意味クラスに属する単語埋め込みベクトルの分布の分析」に関して報告。

【参加者】 Anna Korhonen, Ivan Vulić, Daniela Gerz, Edoardo Ponti, Olga Majewska, Yi Zhu (Cambridge University), Roi Reichart (Israel Institute of Technology)

- LEXICAL プロジェクトのミーティング (2018 年 3 月 21 日)

同ミーティングにおいて「FrameNet と日本語語彙知識の対応付け」に関して報告。

【参加者】 Anna Korhonen, Ehsan Shareghi, Daniela Gerz, Olga Majewska (Cambridge University), Daisuke Kawahara (Kyoto University)