

令和 2 年 4 月 15 日

## 若手研究者海外挑戦プログラム報告書

独立行政法人日本学術振興会 理事長 殿

受付番号 201980396

氏 名 寺西 裕紀  
(氏名は必ず自署すること)

若手研究者海外挑戦プログラムによる派遣を終了しましたので、下記のとおり報告いたします。  
なお、下記記載の内容については相違ありません。

### 記

- 1.派遣先：都市名 テルアビブ (国名 イスラエル)
- 2.研究課題名 (和文) : 長文・複雑な文の構文解析
- 3.派遣期間： 令和 元年 9 月 2 日 ~ 令和 2 年 3 月 20 日 (201日間)
- 4.受入機関名・部局名： Bar-Ilan University, Department of Computer Science

### 5.派遣先で従事した研究内容と研究状況 (1/2 ページ程度を目安に記入すること)

派遣先では自然言語の長文・複雑な文に対する構文解析手法の研究・開発に従事した。語と語の修飾・被修飾関係を同定する依存構造解析のタスクにおいて、科学技術論文や法令文など、難解な専門用語や長く複雑な文が頻出するドメインでは文を高い精度で解析することができない。そこで派遣中に以下四つの研究に取り組んだ。

**(研究 1)** 既存の構文解析器の頑健さについて調査を行った。高精度に解析ができるドメインを対象に、文中の内容語を類義語に置き換えることで構文解析器の振る舞いがどのように変化するか分析した。英文 1,700 文に対して、文中の内容語を一つのみ類義語で置換することで 129,820 の文を得た。これらの文の解析結果をもとの文の解析結果と比較したところ、依存構造木は単語単位で 99.17% 一致し、類義語に対しても一貫して解析ができていたことが明らかになった。

**(研究 2)** 長文では修飾・被修飾語間の距離が長く、依存構造木の階層が深くなる傾向がある。既存の構文解析手法に距離や階層構造などの特徴を明示的に組み入れることで、解析精度がどの程度改善されるか実験を行った。実験の結果、提案手法の解析精度は既存の構文解析手法と同程度にとどまり、階層構造予測のエラー伝搬などに課題が残った。

**(研究 3)** 長文をいくつかのセグメントに分割して解析を行う手法について開発を進めた。セグメントの分割・学習・予測に用いていたコーパスに致命的な欠陥が見つかったため、データの修正を先に進めることに決まった。

**(研究 4)** 長文・複雑な文を多く含む科学技術論文のコーパスの整備を進めた。アノテーションに専門知識が要求されるドメインであるために、文に付与された構文情報に一貫性がなく、また構文情報の欠落などが多く見つかった。これらの問題を半自動で修正する方法の開発を進めた。

## 6.研究成果発表等の見通し及び今後の研究計画の方向性 (1/2 ページ程度を目安に記入すること)

上述の研究1,2を通じて、新聞記事のコーパスで高い性能を発揮する構文解析手法が、科学技術論文のドメインにおいて顕著に誤る事例・傾向について知見を得た。今後は構文解析器が不得手とする言語現象にフォーカスし、研究2,3の手法を取り入れて長文・複雑な文に頑健な構文解析器の研究・開発を引き続き進める。ただし、構文解析器の開発に先立って研究4のデータ整備を進める必要がある。今回の派遣中は研究4のアプローチについてプロトタイプを開発していたところ、新型コロナウイルスの流行の影響により、派遣期間を短縮して帰国するに至った。現在のところ研究成果の発表の見通しは立っていない。

今後は引き続き Bar-Ilan 大学の Yoav Goldberg 教授とともに共同研究を進める予定である。研究計画は以下のとおりである。

**(研究再開後～1年)** コーパス中の非一貫なアノテーションを自動修正し、アノテーションの欠損を半自動で補完する手法を開発する。アノテーションの半自動修正の手法について論文を投稿し、最終的にはオープンソースのツールとして公開を目指す。また、既存の科学技術論文のコーパスに本手法を適用し、本手法により解決されない事例を人手によって修正し、改善されたアノテーションを公開する。アノテーションの高品質化によって構文解析の結果がどの程度改善されたかについて、技術レポートを合わせて提供する。

**(前述の計画達成後～1年)** 上記アノテーションを用いて科学技術論文のドメインにおける既存の構文解析器のエラー傾向を再度精査する。アノテーションの改善により、長文・複雑な文の分解の誤りが低下し、エラー伝搬が低減することで文の部分解析・統合による手法の精度が向上すると期待される。本手法の研究成果を論文・ツールとして公開する。

## 7.本プログラムに採用されたことで得られたこと (1/2 ページ程度を目安に記入すること)

**(研究の場で英語を実践的に用いた経験と自信)** 研究生活では第一に言語の面で苦勞した。イスラエルの公用語はヘブライ語であるが、英語を流暢に話せる人が多く、日常生活の面では全く苦勞しなかった。しかし研究においてはディスカッションやミーティングの場で専門的な内容・意見を正確に理解・伝達する必要がある、日常会話とは異なる英語運用能力が必要であることが分かった。国際会議での発表は事前に原稿を作って練習することができるが、研究活動でのディスカッションは毎回がぶっつけ本番という感覚であり、常日頃から研鑽するしかないと感じた。また同僚がヘブライ語で雑談しているときに、英語で割って入っても仲良くなろうという気概が重要なことも痛感した。自身の研究内容について40人超のメンバーの前で45分のプレゼンテーションをしたことは大きな自信となった。

**(生産性の高い研究機関で活動した経験、コネクション)** Bar-Ilan 大学 NLP 研究室からは毎年トップ会議に面白い論文が何本も通っている。同じ大学院生という立場でありながら何が生産性の違いを生み出しているのか、自分が所属する研究室や自分自身と異なる点について観察し、研究へのより良い取り組み方や姿勢を学ぶことができた。また、こうした環境で切磋琢磨している優れた研究者たちと知り合えたことは、自身の研究者としてのキャリアにとって非常に意義が大きい。

**(異なる文化で働くこと)** イスラエルという、中東に位置するユダヤ教の国で感じたことは多い。まず、ユダヤ教という我々にとって全く馴染みのない宗教が根付いた国での暮らしは驚くことが多かった。平日は日曜日から始まり、週末である金・土曜日は安息日とされ、レストランやスーパーだけでなく公共交通機関など多くの店舗・サービスが休業となる。そのため、否が応でも週末の過ごし方やストレスマネジメントについて考えなければならなかった。また、ガザ地区から飛んでくるロケット弾、アラブ諸国との緊張関係、イスラエル国防軍の兵役など、民族・宗教・歴史・紛争・難民について意識させられる出来事や背景などを肌で感じた。イスラエル生活では聖書を読み、アブラハムの宗教について学び、聖地エルサレムに思いを馳せた。月並みな表現だが、視野が広がり、人生が豊かになった。新型コロナウイルスの流行に対するイスラエル政府の厳格で強制力のある迅速な対応に対して、政府の在り方の違いをまざまざと感じるとともに、外国人として異国で暮らす難しさを思い知らされた。