

# 船井情報科学振興財団第八回留学報告書

大谷直樹

2020 年 12 月

2017 年秋よりカーネギーメロン大学 (Carnegie Mellon University; CMU) の Language Technologies Institute (LTI) に在学中の大谷直樹です。相変わらず外出自粛の毎日ですが、元気に過ごしています。

## 1 近況報告

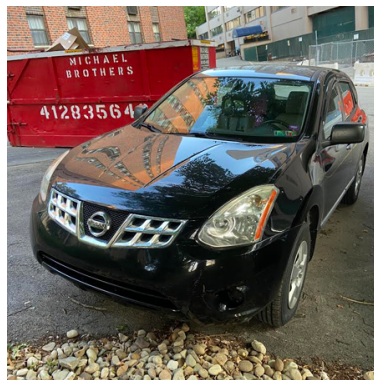
CMU の先輩で奨学生の青木さんがピッツバーグを旅立っていきました。その際に余っていたという日本のカレーのルーをもらったのですが、その美味しさに衝撃を受け、その後しばらく毎週カレーを作っては 3 日くらい食べ続けるという生活を送りました。そこでふと「学業すべてをオンラインで済ませながら自炊で日本食を作り続けるくらいなら、もはや日本に帰ったほうがいいのか」と気づいてしまい、10 月末に帰国しました。

アメリカ国内線の便は満員でしたが、成田までの便はガラガラでした。成田の PCR 検査場のホワイトボードに書いてあった乗客数を見ると、30-40 人程度しか乗っていなかったようです。飛行機の真ん中のキャビンは一切使わず、飲食のサービスも必要最小限でした。入国後 2 週間は公共交通機関を使ってはいけないというルールがあるため、成田から岐阜の実家までは両親が運転する車で運んでもらいました。

日本での生活は最高です。この決断を後押ししてくれた青木さんのカレールーには感謝の気持ちしかありません。帰国後は特に食生活のクオリティーが爆上がりしまして、腹回りがすくすくと成長中です。



(a) CMU にてわたし、青木さん、ファミくん。林くんはこの写真を撮影しながらダブルピースしています。



(b) 車買いました。運転が怖いのもっとばら觀賞用です。



(c) 日本の刺身定食が美味すぎて死にそうです。



Figure 1: オンラインで行われた国際会議の特設サイト。各論文に当てられた個別ページからプレゼンテーション動画と論文が参照できる。会期中はディスカッションフォーラムもついていた。

ミーティングや TA 業務を週に数回早朝にこなさなければいけない点以外はコミュニケーションに不便は感じません。日本でも基本的に家に引きこもっていますが、これはピッツバーグにいたときと変わらないので問題ありません。色々規制が厳しいアメリカよりも社会が動いている感があって (ノーガード戦法とも言う?) 気持ちも明るくなります。

ただ、日本にいながらすべての授業や研究ミーティングに出られるという例外措置がいつまで続くかわからないので、次の学期が始まる前にはアメリカに戻る予定です。それまでにアメリカの感染流行が少しはマシになっていること願っています。

## 2 学業

夏学期中はリモートでインターンをしました。ほとんどの時間は自宅で黙々と作業をするだけで、インターンをしているという実感は正直なところほぼ皆無でした。やはり全く新しい環境で新しく出会う人達に (物理的に) 囲まれながら研究する、という要素がないとかなり寂しいです。

メインの研究テーマである言語理解のための常識的知識の獲得と応用に関しては、相変わらずなかなか芽が出ず、です。今学期は問題設定やデータの質を見直しました。狙っていることがやや複雑なので、良い計算モデルを考え出す以前に正確なデータをできるだけ簡単に多く取るという点で苦労しています。

メインプロジェクトとは別に以前から行っていた研究について書いた論文を、11月の学会で研究発表をしました<sup>1</sup>。この論文で取り組んだのは多言語単語ベクトルと呼ばれる技術で、大量のテキストと辞書などの翻訳資源を使って、複数の言語の単語を機械が扱いやすい実数ベクトルで表すものです。多言語処理技術の入力コンポーネントとして広く使われています。私が着目したのは “New York-ニューヨーク” や give up-諦める”といった、複数の単語で構成された表現と一単語の対応です。前者は「単語」ではないため、単語ベクトルに関する既存研究では無視されてきました。私の論文はこの点を指摘し、データの前処理を少し工夫するだけでその問題が解消されることを示しました。

国際会議は完全にオンラインでした。プレゼンテーションは事前に録画し、SlidesLive<sup>2</sup>というサービ

<sup>1</sup>Naoki Otani, Satoru Ozaki, Xingyuan Zhao, Yucen Li, Micael St Johns and Lori Levin. 2020. Pre-tokenization of Multi-word Expressions in Cross-lingual Word Embeddings In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pages 4451–4464, November. Association for Computational Linguistics.

<sup>2</sup><https://slideslive.com/>

スを使ってアップロードします。会期中は会議の特設 Web サイトからすべての論文のプレゼンテーションが見られるようになっており (図 1)、発表者のタイムゾーンに合わせて行われる QA セッションでリアルタイムにコミュニケーションができます。Gather.town という Web サービスを介して、仮想空間を動き回りながら人と会話するという仕組みも導入されていました。仮想空間上で他人と近づくと自動的に音声とカメラ映像がオンになって、あたかも現実の会場で歩き回りながら人と会うような経験ができる、というものです。新鮮で面白かったです。

今学期の研究以外の活動としては、LTI の招待講演シリーズ<sup>3</sup>のティーチングアシスタントをしました。TA と言っても学生の質問に答えたり試験問題の採点をしたりといった仕事はなく、事務的な作業がメインです。具体的には

- 講演者と連絡をとって学生との個別ミーティングを設定する
- 講義中に出席確認用の簡単なクイズをつくる
- 講義を録画して、リアルタイムで出席できない学生向けに公開する
- 録画した動画をいい感じに編集して YouTube にアップロードする<sup>4</sup>
- 学期末にクイズの結果を集計して成績表を出す

というようなお仕事をしていました。日本に帰国後は午前 3 時から始まる講義に参加しながらクイズをつくるのが辛かったです。

### 3 おわりに

今学期の後半は久々の日本での生活で (特に食事が) 幸せでした。年末年始は美味しい料理と日本酒で 100 キロくらい太りたいと思います。1 月半ばにはいま色々な意味で世界をリードする国アメリカに再び戻る予定ですが、引き続き健康に気をつけてがんばります。

---

<sup>3</sup><https://lti.cs.cmu.edu/lti-colloquium>

<sup>4</sup>LTI の YouTube チャンネル (<https://www.youtube.com/c/LTIatCMU/playlists>) 『Colloquium』から見られます。