

船井情報科学振興財団第四回留学報告書

大谷直樹

2018 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) 修士課程に在学中の大谷直樹です。今学期の最後は期末試験・レポート、論文、博士課程への申請が重なり、寝ても締切に追われる夢を見るような有様でしたが、なんとかすべて片付けました。

1 近況報告

1 年半前に CMU に来たときからすでに工事が始まっていたビジネススクールの新校舎が夏にとうとう完成したり、キャンパスの歩道がきれいになったりしました。大都会ピッツバーグでの生活は毎日がエキサイティングなことばかりで、それ以外に何も思い出せないほどです。

今学期の最大のイベントは何と言っても 2 年ぶりの博士課程申請でした。私は現在 LTI の修士課程に在籍していて、博士課程に上がるためにはもう一度外部から受けるのと同様の書類を揃え、入試に合格しなければなりません。とはいえ、LTI の教員から推薦状をもらえるという大きなアドバンテージがあり、内部進学の高め (のはず) です。ただ、友達や先生方の話を聞いていると、教員の懐事情がかなり重要なファクターだという気がしています。昨年度はトップ会議に複数本の論文を通していた先輩学生が (おそらくファンディングの問題で) 不合格でした。私はやれるだけのことはやったつもりですが、果たして結果がどうなるか、祈るばかりです。

私生活ではビールを作るなどしました (下図)。

2 講義

今学期は 4 つの講義を履修しました。学期中は常に宿題とコースプロジェクト (+ 自分の研究) をこなすことに精一杯で心の余裕がありませんでした。講義を楽しみながら消化できるのは 3 つが限界かもしれません。

Introduction to Machine Learning: LTI 修士課程 (他の課程も?) の学生のほとんどが履修するよう
な人気講義で、履修者はトータルで 300 人くらいはいたのではないかと思います。TA も 8 人くらいいて、これまで履修した中で最も大きな講義でした。4 回の宿題、2 回の筆記試験、コースプロジェクト、という全部盛りな内容で、かなりの QOL を持っていられました。それぞれの質自体は残



(a) CMU のキャンパス。今年のピッツバーグには秋が 5 日くらいしかなかった気がします。半旗になっているのは (父) ブッシュ元大統領の逝去を受けて。



(b) ビールづくりの風景。鍋の中には大麦の茹で汁が入っていて、これから糖を投入するところ。かき混ぜ棒を持っている緑色の生き物はビールの精。



(c) ビールを飲むビールの精。

念ながら期待外れでした。宿題は Piazza という掲示板のやり取りで補完しなければ解けないような不完全な問題が多く、コースプロジェクトは学生にやらせっぱなしな雰囲気が強くて意義を感じませんでした。精神力は鍛えられました。

Grammars and Lexicons: 言語学のうち、語の成り立ちと文法に焦点を当てたコースです。私は恥ずかしながら言語の処理を専門にしながら言語学の基礎を真面目に勉強したことがなかったので、この講義から学ぶことが多くありました。講義では英語の文法を主に取り扱うため、ついていくのがとこどころ大変でした。

Linguistics Lab: 先生の指導を受けながら自由なトピックについて研究を進める講義です。今回はたまたま受講者が私しかおらず、私が興味を持っていた単語の定義 (後述) に取り組みました。言語学出身の先生とのディスカッションは毎回とても刺激的でした。また、研究経験を積みたいという学部生と高校生 (1)¹ がチームに加わって、二人に指示を与えながら研究を進めるという貴重な経験をしました。このトピックはまだまだ掘り下げられそうなので、来学期もこの研究を続けたいと思っています。

LTI Colloquium: 前学期と同じく、毎週一人、自然言語処理分野で著名な研究者が招かれる招待講演です。今学期はスピーカーとランチをしたり、マンツーマンのミーティングをしたりできる機会に3回参加しました。

3 研究

8月、自然言語処理分野で有名な COLING という国際会議に参加しました。この会議は他の国際会議に比べて言語学者が比較的多く参加するのが特徴です。場所は Santa Fe。アドビ様式と呼ばれる赤土で作られた建物が特徴的な街でした。私は (いつも通り) 連日お酒の飲み過ぎで具合が悪くなりながら、発表をしたり聞いたり、人と喋ったりしました。この会議を通して言語学への興味が強まり、今学期は言語学寄りの講義と研究に取り組みました。そういう意味で、この会議は今年の一つの転機になりました。

前述の通り、今学期は Linguistics Lab という講義で「単語の定義」を研究しました。単語は言語処理技術の根幹に関わる重要な概念です。自然言語処理も工学なので、他の工学分野と同様、(1) 入力を受け取って、(2) 何らかの処理をして、(3) 求める出力を得るという流れが基本です。さて、その入力は何かと言うと、多くの場合、文、すなわち単語の列です。では単語とは何なのでしょう？英語のように空白で書き言葉を区切る言語では、空白で切られたものを単語と考えられそうです²。一般的には、その上に縮約形を分解する操作を加えます。例えば、“I’m”を“I”と“m”に、“isn’t”を“is”と“n’t”に分けます。これが言語処理で使われる「英語の単語」です。いま言語処理においてステップ(2)の主流になっている方法は、この「単語」一つひとつに実数ベクトルを対応させて、そのベクトル同士の演算によって言語の構造や意味に関する処理を行うというものです。ここで問題となるのが、その「単語」が果たして言語構造や意味の構成物としてふさわしいものになっているか、ということです。

実際、空白区切りによる「単語」が意味の構成物に対応しない例は簡単に見つかります。例えば、hot dog は hot な dog ではありません。このような構成性 (compositionality) が崩れるような複数語からなる表現は multi-word expressions (MWEs) と呼ばれ、多くの言語に現れます。今学期はこの MWE をひとつのまとまりにして後段のステップ(2-3)を行った場合にどのような変化が現れるかを研究しました。

もうひとつ並行で行っている研究は感情分析に関わるものです。感情分析というのは、ざっくり言うと文の書き手が特定の対象について肯定的な感情を持っているか、否定的な感情を持っているかを分析するタスクです。肯定的か否定的かを当てるとするのが標準的な問題設定ですが、私はさらに「なぜ」肯定的/否定的なのかを分析するという課題を解こうとしています。先程示した3つのステップでいうと、(3)出力ステップにあたる研究です。今学期は心理学関係の文献を調べたりして出力のデザインについて考えました。いまデータセットを作っていて、これを使った分析を今後行うつもりです。

¹アメリカのトップ大学の入試では研究経験が非常に効くそうです。この高校生は無事にスタンフォード大学に合格しました。

²日本語や中国語のように空白区切りを使わない言語はより多くの議論が必要です。

4 中国人の妻をアメリカに連れてくる

前回の報告書で、中国人の妻をアメリカに呼ぶための F2 ビザについて触れました。もう一年も前のことですが、誰かのお役に立てるかもしれないのでここに書き留めておきます。³

F2 ビザというのは学生の配偶者に与えられるビザです。就学・就職に厳しい制限がありますが、F1 ビザ (学生ビザ) 持ちの学生の配偶者であれば比較的簡単に取得することができます。ただし、配偶者が外国人の場合は結婚状態の証明に手間がかかるのでなかなか油断できません。

- (1) 入籍する: 私は日本の市役所で入籍しました。「結婚具備証明書」という双方が結婚に必要な条件 (未婚であること、親族でないこと、等) を満たしている書類の準備が最大の関門でした。相手国や自治体によってその書類の書式が異なることに加えて、市役所の方が不慣れな可能性があるため、審査に不確実性があります。私は母にも手伝ってもらって事前に市役所に書類の要件を問い合わせたので、幸い一発でクリアできました。
- (2) 妻の中国側の戸籍情報を変更する: 日本の市役所に婚姻届を提出するだけでは中国の戸籍上の婚姻情報が変わりません。これを変えるためには、日本で結婚したという証明 (婚姻届受理証明書) を妻の実家がある地区の戸籍管理部に提出する必要があります。市役所で発行できる証明書はそのままでは中国で効力を持たないため、認証を日本側 (外務省など) と中国側 (在日中国領事館) で受けます。どちらの認証もただのシールのものなのですが、なぜかその場では処理してもらえず後日郵送されてきました。おそらくシールを貼った書類の周りで踊るなどといった儀式をしていたのでしょう。認証済みの婚姻届受理証明書 (+ 中国語の翻訳) に私のパスポートのコピーなどを合わせて中国の戸籍管理部に提出すると、無事に妻の戸籍を書き換えることができました。
- (3) F2 ビザ申請用の書類を集める: 認証済みの婚姻届受理証明書に加えて、日本人の配偶者に F2 ビザを発行する場合と同様の書類を集めます。当時、妻は中国、私はアメリカにいたので、F2 ビザ取得に必要な I-20 を学校 (CMU) に発行してもらい、生活費証明を中国にいる妻のもとに送りました。妻の方では戸籍のコピーなど、一般的に F2 ビザで必要とされる書類を集めました。
- (4) 妻に面接に行ってもらう: F1 ビザと同様に F2 ビザの発行にも簡単な面接があります。これ以上私にできることはないため、遠くから応援するのみです。面接が終われば晴れて F2 ビザをゲットです。ちなみに妻によると面接自体はあっさり終わったそうですが、書類のチェックに結構待たされたそうです。中国の方は一年ごとのビザしか発行してもらえないことが多いみたいですが、妻は私と同じ有効期間のビザを出してもらえました。

こうしたステップを乗り越えて、なんとか妻をアメリカに呼ぶことができました。やはり妻は精神的な支えになってくれますし、一緒にいるほうがなんだか生活に張りがある気がします。来てもらえてよかったです。

5 おわりに

今学期は興味のある研究トピックに取り組む機会に恵まれ、とても満足しています。次の半年には博士課程申請の結果が出て夏からの居場所が決まったり、トップ会議の論文投稿締切が複数個あったり、また忙しくなりそうです。健康に気をつけて、次の学期も充実したものになるよう頑張ります。

³ここに書かれた情報はあくまで一例に過ぎないことに注意してください。