

2. 現在までの研究状況 (図表を含めてもよいので、わかりやすく記述してください。様式の変更・追加は不可(以下同様))

- ① これまでの研究の背景、問題点、解決策、研究目的、研究方法、特色と独創的な点について当該分野の重要文献を挙げて記述してください。
- ② 申請者のこれまでの研究経過及び得られた結果について、問題点を含め①に記載したことと関連づけて説明してください。
なお、これまでの研究結果を論文あるいは学会等で発表している場合には、申請者が担当した部分を明らかにして、それらの内容を記述してください。

【これまでの研究背景】

計算機に日本語などの言語を理解させる上で、その文の構造を把握すること(構文解析)は重要な役割を果たす。**組み合わせ範疇文法(CCG)[1]**に基づく構文解析は、他の言語理論で扱いが困難な並列句などの言語構造を適切に表現できるという言語学的な強みに加え、文構造から文の論理表現への変換を簡単にできるという特徴があるため、与えられた2文が論理的に含意するかを予測する含意関係認識や言語によるロボット制御など多くの自然言語処理タスクで利用されている[2]。

CCGの特徴は、**カテゴリ**が単語の詳細な統語的性質を記述する一方で文法規則は少数からなることであり、この理論に基づく構文解析では各単語に対する**カテゴリ**が決まれば文構造の大部分が決定する。図1の動詞 had のカテゴリは主語と目的語を取って文(カテゴリ S)を導出するという制約を表し、この文の構造は図のカテゴリ列のもとで一意である。この特徴を利用して、Lewisらにより**CCG木の確率を、各語に付与されるカテゴリの確率の積で表す確率モデルが提案された[3]**。木の確率の定義の単純さから、CKY法を拡張したA*解析を用いることで高速に最適解を求めることができる一方で、カテゴリの予測をニューラルネットで行うことで、当時の最も高い構文解析精度を示した[4]。

【問題点】

Lewisらの手法により高速、高精度なCCG解析が可能となったが、彼らのモデルにはその**単純さに起因する問題**が存在する。彼らのモデルでは木構造の確率はカテゴリのみに依存しているが、**同じカテゴリ列(よって同じ確率)から複数の木が得られる場合があり、そのときどれを解析結果とすべきか決定できない**。例えば、“house in suburbs of Paris”に対して、各語に正しいカテゴリを付与しても図2の2つの構造があり得る。彼らのこの問題への対処法は、変換ルールを用いてCCG木から単語間の係り受け関係を表す依存構造(図2の有向エッジの集合)に変換し、依存構造においてより長い依存エッジを伴う木を解析結果として選ぶという(経験的にうまくいく)方法である。しかし、これによれば長距離エッジを伴う(a)が解析結果となるが、「パリ郊外の家」という解釈に対応するのは“of Paris”が“suburbs”に係る(b)であり、この例では誤った予測をしてしまう。このように、この対処法は過度に単純であり一般的にうまく機能しない。

【解決策と研究目的・方法】

① 依存構造を考慮する新しいモデルの提案

申請者は、上述の問題を解決するために、**CCG木の確率を各単語のカテゴリ列と依存構造の確率の積とする新しい確率モデルを提案した**。このモデルでは、カテゴリ列だけでは等確率になってしまう木構造の問題に対し、依存構造をルールのために利用するのではなく、CCG木の確率モデルに取り込むことで対処する。よって先の問題はどちらの依存構造が尤もらしいか、という観点で柔軟に対処される。また、このモデルでもA*探索に必要なヒューリスティック値が定義でき、A*解析で高速に最適解を計算できる。

② カテゴリ列と依存構造を同時に予測するニューラルモデル

この手法がうまく機能するためには、依存構造の予測が高精度にできなければならない。従来の依存構造解析はCKY法などのアルゴリズムを用いていたが、近年、ニューラルネットのみを用いて高精度かつ効率よく依存構造を予測する手法が示されている[5]。申請者は、その手法を拡張し、**単語のカテゴリと依存構造を同時に予測するニューラルネットの構造を提案した**。これにより、依存構造の確率を用いても効率的に解析可能である。

カテゴリ: $(S \setminus N) / N$

主語、目的語(N)を取り文(M)を導出

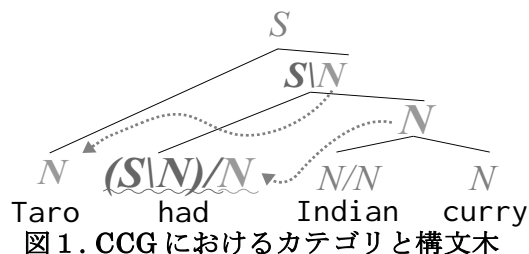


図1. CCGにおけるカテゴリと構文木

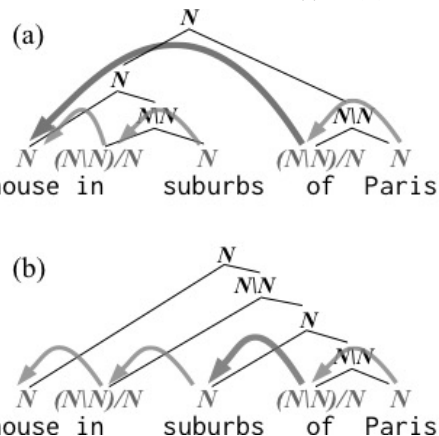


図2. 同じカテゴリ列から導出されるCCG木とその依存構造

[1] M. Steedman, “The Syntactic Process”, 2000
 [2] Y. Artzi and L. Zettlemoyer, “Weakly Supervised Learning of Semantic Parsers for Mapping Instructions to Actions”, 2013
 [3] M. Lewis and M. Steedman, “A* CCG Parsing with a Supertag-factored Model”, 2014
 [4] M. Lewis et al., “LSTM CCG Parsing”, 2016
 [5] T. Dozat and C. D. Manning, “Deep Biaffine Attention for Neural Dependency Parsing”, 2016

(現在までの研究状況の続き)

【特色と独創的な点】

この研究の独創的な点は、比較的困難な CCG 解析に、高精度かつ効率よく予測できる依存構造を援用した点である。また、依存構造を用いることは**解析の高速さという強み**がある。Lewis らの手法に対し、カテゴリ以上の構造を考慮する拡張を行った研究は Lee らによるものもある[6]。彼らのアプローチは、複雑なニューラルネットを用いて CCG 木の尤もらしさを直接的に評価する。一方で提案法では CCG 木の背後の依存構造を用いて間接的に構造の良さを考慮する。依存構造は図 2 のそれぞれのエッジに分解して考えることができるため、CCG 木の確率を、各単語のカテゴリとそれぞれの依存エッジの確率という局所的な要素の積に分解できる。これらは A*探索を行う前にニューラルモデルで効率よく計算でき、これにより Lee らのものと比べ**約1.6倍高速**に構文解析を行うことができる。

【研究経過及び得られた結果】

英語と日本語の CCGBank コーパスを用いて、提案モデルの学習と性能評価の実験を行なった。

英語の実験では、提案法は既存の Lewis ら、Lee らの手法より高い**88.8%**という精度を示し、英語の CCG 解析において**最高精度**となった。

	Lewisら	Leeら	提案法
精度	88.0%	88.7%	88.8%

英語CCGBankでの実験結果

	既存	Lewisら	提案法
精度	87.5%	81.5%	91.5%

日本語CCGBankでの実験結果

日本語でも提案法が既存手法に対して最も高い解析精度を示し、構造予測の正答率においてこれまで最高精度であった手法に対し**4.0%改善**し、Lewis らの手法を単純に日本語に適用したものには**10.0%の改善**を示した。Lewis らの手法を日本語に適用した場合には解析精度はかなり悪く、日本語の解析では依存構造を考慮することの有効性が明らかになった。日本語では、カテゴリの情報だけでは文構造が曖昧な場合が多く、その原因としては比較的自由的な語順が挙げられる。

これまでの研究は、自然言語処理学会、自然言語処理研究会で発表し[業績 4,5]、自然言語処理学会では優秀賞を受賞した[業績 7]。また自然言語処理における最難関会議の ACL2017 で発表予定である[業績 2]。申請者は、構文解析器の実装、実験、原稿の執筆まで本研究を主導的に執り行った。

[6] K. Lee et al., “Global Neural CCG Parsing with Optimality Guarantees”, 2016

3. これからの研究計画

(1) 研究の背景

2. で述べた研究状況を踏まえ、これからの研究計画の背景、問題点、解決すべき点、着想に至った経緯等について参考文献を挙げて記入してください。

【研究の背景】構文解析において解析が困難なことが知られている表現に**並列句**がある。並列句とは次のような and や or と、それにより接続される 2 つの句からなる表現である。

(a) I cooked and might eat the beans.

この例のように and 等の語でどの表現が並列されているかを解析すること (**並列句の範囲同定**) は、文の意味を理解する上で重要である。しかしこの問題は難しいことが知られ、広く利用されている依存構造解析器の Stanford Parser [1] を用いて (a) の文を解析すると “cooked” と “might eat the beans” が並列しているという誤った予測をし、また並列句の範囲同定の専用システム [2] でさえも、(a) のように並列されている句のペアの予測において 72.7% の F 値の精度にとどまっている。

【問題点・解決すべき点】並列句の範囲同定には、**統語的な特徴**と**意味的な特徴**の両方を同時に考慮する必要がある。例えば、上の (a) の例では cooked が他動詞として機能する確率が高いことが考慮できれば、Stanford Parser の場合のような誤った解析は防ぐことができる。一方で意味を考慮することで、どの句と句が並列されているかを解消できる場合がある。

(b) The broadcast and publishing company

この例では、“The broadcast” と “publishing company” のペアの場合よりも “broadcast” と “publishing” のペアのほうが意味的に近いという知識が範囲同定に有効である。そこで、統語的特徴を考慮するために CCG を用い、意味的な特徴を捉えるために、単語や句の類似度を計算する上で強力なニューラルネットを利用することが考えられる。

【着想に至った経緯】後述するように CCG の強みの 1 つとして、**並列句の記述能力が高い**ことである。しかしこれまで並列句の範囲同定の課題に CCG を用いた手法はない。この理由の 1 つとして依存解析など他の構造予測と比較すると CCG による構文解析が困難であったからであると考えられる。申請者のこれまでの研究による CCG 解析の高精度化、高速化によって CCG を用いたアプローチも可能になったと考える。

[1] D. Chen and C. D. Manning, “A Fast and Accurate Dependency Parser using Neural Networks”, 2014

[2] J. Fidler and Y. Goldberg, “A Neural Network for Coordination Boundary Prediction”, 2016

(2) 研究目的・内容 (図表を含めてもよいので、わかりやすく記述してください。)

- ① 研究目的、研究方法、研究内容について記述してください。
- ② どのような計画で、何を、どこまで明らかにしようとするのか、具体的に記入してください。
- ③ 所属研究室の研究との関連において、申請者が担当する部分を明らかにしてください。
- ④ 研究計画の期間中に異なった研究機関（外国の研究機関等を含む。）において研究に従事することを予定している場合はその旨を記載してください。

【研究目的】

本研究は、並列句の構造を高精度に予測できる CCG に基づく構文解析手法の開発を目的とする。

【研究方法・研究内容】

これまで申請者が研究してきた CCG 解析の手法を拡張し、上述の並列句の**統語的、意味的な特徴**を同時に考慮しながら、その範囲同定を高精度に行う手法を構築する。

① CCG による統語的な制約の表現

図 3 のように、CCG の文法規則では同じカテゴリを持つ句のみ並列できるとされている。このような分析は適切な粒度を持つ CCG のカテゴリによって実現される。図の例では、まず“cooked”と“might eat”の句が結合してから共通の目的語“the beans”をとる、という統語的な制約を表現する。一方で、同一の文法規則が、以下のような他の文法理論では扱いが困難な並列句の構造も説明できるため、CCG はこの問題を扱うのに最適な文法理論である。

例. Give (a teacher an apple) and (a policeman a flower).

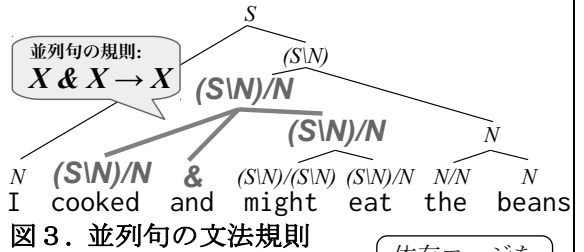


図 3. 並列句の文法規則

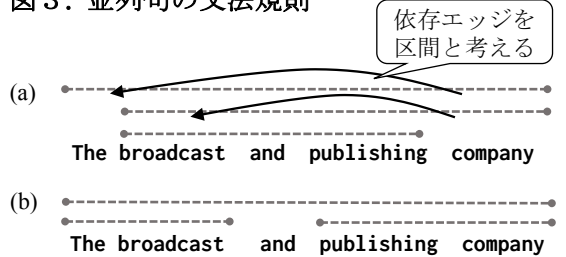


図 4. 依存エッジの区間としての解釈

② ニューラルネットを用いた並列された句の類似度学習

これまでの申請者の研究で依存構造を用いたが、そのエッジは向きを無視して無向エッジと再解釈することができる。このとき、エッジに伴う確率値は、エッジが貼る区間内の**単語列が自然な句を成しているか**を表すと解釈できる(図 4)。これまでのカテゴリ列と依存構造による CCG 木モデルをカテゴリ列とそのような区間の集合によるモデルと考え、図 4 (a)の“broadcast and publishing”の句が(b)の並列句より意味的に自然であることを、それらの区間の確率値を通して直接的にモデル化する方策が建てられる。並列句の自然さは、並列された2つの句が意味的に似ていればより自然と考えられる。ニューラルネットを用いた自然言語処理では、**単語を実数値ベクトル**で表現することで単語や句の類似度計算に強力であるため、この特徴を並列句の自然さの計算に利用する。

【どのような計画で、何を、どこまで明らかにするか】

研究 A 並列句範囲同定の実験データ作成

CCG 木を文に付与した CCGBank コーパスは、英語の新聞記事に句構造を付与した Penn Treebank コーパスからルールにより変換して得られたものである。Penn Treebank における並列句のアノテーションは間違いが多いことがよく知られており、並列句の範囲同定の実験のために、誤った箇所を修正した拡張版が作られている[1]。そのため、本研究で CCG 解析による並列句の範囲同定に取り組むために、その拡張版に基づき並列句のアノテーションを修正した新たな CCGBank を作成する。

研究 B 依存エッジの再解釈によるモデルの拡張

依存エッジの確率を単語列の句としての自然さを表すと解釈することで、これまでの申請者による研究の CCG 解析手法を拡張する。具体的には、**単語列の自然さをより直接的にモデル化するニューラルネット構造を構築する**。また、並列句に関する既存研究では、並列句の以下の2つの特徴を捉えることが有効であることが示されている。

- 1. 対称性 (a teacher an apple と a policeman a flower の意味と構造における対称性)
- 2. 置換可能性 (Give a policeman a flower のように並列句の片側で全体を置換しても自然なこと)

提案モデルでは並列句の類似度に加え、これらを捉えられるネットワーク構造と素性の設計を行う。

研究 C 外部タスクを用いた CCG 解析器の並列句範囲同定の性能評価

上に述べたように、並列句の範囲を適切に当てることは、文の意味解釈に大きく影響する。そのため、本研究の取り組む課題は、与えられた2文が含意の関係にあるか予測する含意関係認識や質疑応答など意味に関わるタスクと密接な関係にある。そこで、**本研究で開発する並列句の範囲同定手法による性能改善をこれら外部のタスクで評価する**。具体的に、CCG 解析器を用いてこれらのタスクの入力文を解析し、その解析結果の文構造から変換して得られる論理式を用いてこれらのタスクの推論に取り組む。そのときに**研究 B**で行った拡張による影響をこれらのタスクで評価する。

[1] J. Fidler and Y. Goldberg, “Coordination Annotation Extension in the Penn Tree Bank”, 2016

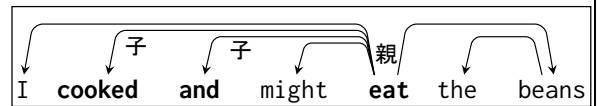
(3) 研究の特色・独創的な点

次の項目について記載してください。

- ① これまでの先行研究等があれば、それらと比較して、本研究の特色、着眼点、独創的な点
- ② 国内外の関連する研究の中での当該研究の位置づけ、意義
- ③ 本研究が完成したとき予想されるインパクト及び将来の見通し

【先行研究との比較】既存研究の多くは、並列句の範囲同定専用の手法である[1]。本研究は、この問題は文構造解析の一部であるという考えから構文解析のなかで解決する。これにより並列句を含む文全体の構造が手に入るため、応用しやすいという強みがある。また、並列句のパターンや解析対象の分野を限定して並列句の解析に取り組む研究も多い[2]。本研究は、並列句に強い CCG 文法と句の自然さという一般性の高い方策に基づくため、そのような限定は必要ない。本研究に近い手法として、HPSG 文法の制約と文字列の類似度を利用した手法がある[3]。この手法は、1つの文を処理するために2つの解析器を複数回実行するため、実行速度の問題がある。提案法は一度 A*探索を行うだけであり高速である。

【当該研究の位置づけ、意義】自然言語処理では、依存構造が広く研究され、利用されているが、並列句の取扱には問題が多い。精力的に推進されている **Universal Dependencies**



コーパスでは、右図のように and で並列された右の語を常に並列句全体の親として分析する。しかし、この分析では図の2つの動詞が目的語を共有している事実を表現できない。CCG では、図4のようにこの事実を表現でき、これは論理表現にも反映される。

【インパクト及び将来の見通し】並列句は広範に用いられる言語表現であるが、元来その構造の解析は困難なことが知られていた。申請者のこれまでの研究で、この構造の取扱に強力な CCG に基づく構文解析器の準備が整っている。この構造の解析は多くの自然言語処理の課題の基礎であるため、完成した暁にはそれら多くの課題の精度向上に貢献することができる。

[1] J. Ficler and Y. Goldberg, “A Neural Network for Coordination Boundary Prediction”, 2016

[2] 山腰貴大 et al., “ニューラル言語モデルを用いた法令文の並列構造解析”, 2017

[3] A. Hanamoto et al., “Coordination Structure Analysis using Dual Decomposition”, 2010

(4) 年次計画

申請時点から採用までの準備状況を踏まえ、DC1 申請者は1～3年目、DC2 申請者は1～2年目について、年次毎に記載してください。元の枠に収まっていれば、年次毎の配分は変更して構いません。

(申請時点から採用までの準備)

CCG 木を可視化、操作する GUI ツールの開発

拡張版 Penn Treebank から変換を行って CCG のコーパスを作ることは困難であることが予想される。できるだけこの過程を平易化するため、木を可視化し、その上で操作できる GUI ツールを開発する。

(1年目)

拡張版 Penn Treebank から CCG 木への変換

- 性能評価のために並列句の範囲についての情報を抽出
 - 現時点の CCG 解析器がどの程度、並列句を正しく予測できるか評価
- Penn Treebank の並列句を含む構文木を主な対象として CCG 木への変換アルゴリズムを設計
- 先に作った GUI ツールを用いて変換して得た CCG 木を評価、修正
- データの作成と、現時点での予測精度についてまとめて国際学会で発表

(2年目)

CCG 構文解析器に並列句の区間を学習させる

- ニューラルネットを用いた句の類似度の計算手法
 - 対称性や置換可能性を捉える方法を検討
- CCG 構文解析器の開発
 - これまでに作成した CCG 解析器をもとに拡張
 - ニューラルネットモデルの学習に1年目で開発したコーパスを利用
- 提案モデルの性能を評価し、国際学会で発表
 - 1年目で開発したコーパスで性能評価
 - 含意関係認識、質疑応答などの外部タスクでの性能評価

(3年目) (DC2 申請者は記入しないでください。)

(年次計画の続き)

(5) 人権の保護及び法令等の遵守への対応

本欄には、研究計画を遂行するにあたって、相手方の同意・協力を必要とする研究、個人情報の取り扱いの配慮を必要とする研究、生命倫理・安全対策に対する取組を必要とする研究など法令等に基づく手続きが必要な研究が含まれている場合に、どのような対策と措置を講じるのか記述してください。例えば、**個人情報を伴うアンケート調査・インタビュー調査、国内外の文化遺産の調査等、提供を受けた試料の使用、侵襲性を伴う研究、ヒト遺伝子解析研究、遺伝子組換え実験、動物実験など**、研究機関内外の情報委員会や倫理委員会等における承認手続きが必要となる調査・研究・実験などが対象となりますので手続きの状況も具体的に記述してください。

なお、該当しない場合には、その旨記述してください。

本研究で用いる実験データはすべて公開されているコーパスであるため、該当しない。

4. 研究業績（下記の項目について申請者が中心的な役割を果たしたもののみ項目に区分して記載してください。その際、通し番号を付すこととし、該当がない項目は「なし」と記載してください。申請者にアンダーラインを付してください。業績が多くて記載しきれない場合には、主要なものを抜粋し、各項目の最後に「他〇報」等と記載してください。査読中・投稿中のものは除く）

(1) 学術雑誌等（紀要・論文集等も含む）に発表した論文、著書（査読の有無を区分して記載してください。査読のある場合、印刷済及び採録決定済のものに限ります。）

著者（申請者を含む全員の氏名（最大 20 名程度）を、論文と同一の順番で記載してください。）、題名、掲載誌名、発行所、巻号、pp 開始頁—最終頁、発行年をこの順で記入してください。

(2) 学術雑誌等又は商業誌における解説、総説

(3) 国際会議における発表（口頭・ポスターの別、査読の有無を区分して記載してください。）

著者（申請者を含む全員の氏名（最大 20 名程度）を、論文等と同一の順番で記載してください。）、題名、発表した学会名、論文等の番号、場所、月・年を記載してください。発表者に〇印を付してください。（発表予定のものは除く。ただし、発表申し込みが受理されたものは記載しても構いません。）

(4) 国内学会・シンポジウム等における発表

(3)と同様に記載してください。

(5) 特許等（申請中、公開中、取得を明記してください。ただし、申請中のもので詳細を記述できない場合は概要のみの記述で構いません。）

(6) その他（受賞歴等）

(1) 学術雑誌（紀要・論文集等も含む）に発表した論文、著書

なし

(2) 学術雑誌等又は商業誌における解説、総説

なし

(3) 国際会議における発表（査読あり）

1. ○Masashi Yoshikawa*, Hiroyuki Shindo**, Yuji Matsumoto***, “Joint Transition-based Dependency Parsing and Disfluency Detection for Automatic Speech Recognition Texts”, In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2016), pp. 1036-1041, Austin, Texas, USA, November 2, 2016
2. ○Masashi Yoshikawa*, Hiroshi Noji**, Yuji Matsumoto***, “A* CCG Parsing with a Supertag and Dependency Factored Model”, In Proceedings of Conference of 55th Annual Meeting of Association for Computational Linguistics (ACL-2017), Vancouver, Canada, August 1, 2017

(4) 国内学会・シンポジウム等における発表

【ポスター・査読なし】

3. ○吉川将司*, 進藤裕之**, 松本裕治***, 「話し言葉の依存構造解析における音声認識誤りの影響の評価」, 言語処理学会 第 22 回年次大会 発表論文集, pp477-480, 宮城, 2016 年 3 月
4. ○吉川将司*, 能地宏**, 松本裕治***, 「係り受け構造との同時予測による A* CCG 解析」, 言語処理学会 第 23 回年次大会 発表論文集, pp274-277, 筑波大学, 2017 年 3 月

【口頭・査読なし】

5. ○吉川将司*, 能地宏**, 松本裕治***, 「係り受け構造とスーパータグの同時予測による A* CCG 解析」, 情報処理学会 第 231 回自然言語処理研究会・第 116 回音声言語情報処理研究会 合同研究発表会, 大阪大学, 2017 年 5 月

(5) 特許等

なし

(6) その他

6. NTCIR-12 MedNLPDoc 8 チーム中 6 位
澤井裕一郎*, 大村舞*, 大内啓樹*, 永井優城*, 吉川将司*, 進藤裕之**, 山田育矢†, “Similarity Matrix Model for the NTCIR-12 MedNLPDoc Task”
7. 言語処理学会 第 23 回年次大会 優秀賞 受賞 2017 年 3 月（受賞対象文献(4)-4）
8. ソフトウェアの公開 日本語、英語 CCG 構文解析器 depccg
<https://github.com/masashi-y/depccg>
9. 平成 29 年度奈良先端科学技術大学院大学優秀学生（29 年度学費全額免除）

* 奈良先端科学技術大学院大学情報科学研究科-学生

** 奈良先端科学技術大学院大学情報科学研究科-助教

*** 奈良先端科学技術大学院大学情報科学研究科-教授

† 株式会社 Studio Ousia-代表取締役 CTO

申請者登録名 吉川将司

5. 自己評価

日本学術振興会特別研究員制度は、我が国の学術研究の将来を担う創造性に富んだ研究者の養成・確保に資することを目的としています。この目的に鑑み、申請者本人による自己評価を次の項目毎に記入してください。

- ① 研究職を志望する動機、目指す研究者像、自己の長所等
- ② 自己評価する上で、特に重要と思われる事項（特に優れた学業成績、受賞歴、飛び級入学、留学経験、特色ある学外活動など）

①【研究職を志望する動機】

高校生の頃にはじめてフランス語を勉強して以来、世界の言語にとっても興味を持った。特に惹かれたことは、語順や語の構成の仕方など、一見したところ全く異なっている世界の言語の間に必ず存在するに違いない共通性についてである。そのことに対する探究心から学部時代には主に外国語学と言語学に専念した。学部後半で言語学に対する数理的なアプローチや論理学に魅力を感じ、形式的な手法による言語の記述や機械学習を用いて計算機に言語を理解させるということに興味を持ち、現在の研究室の門を叩いた。現在は、計算機の能力と機械学習理論が大きく発展していく時代のなかで、今後どのように、どのようなことが言語について明らかになっていくのかということの期待感と、そのような時代に最先端の言語研究に携われる喜びを感じ研究職を志望する。

【目指す研究者像】

文系・理系といった区別を取り払った学際的な研究者になりたいと考える。 今後の数十年において言語の研究は大きく前進するだろうと考える。それは、従来の言語学に加え、大量のデータから知識を取り出す機械学習的手法や、心理学や神経科学の発展により、多方面から言語についての新事実が明らかになっていっているからである。そのようなときに常に言語研究の最前線にキャッチアップし自身も参加して研究していきたいと考えている。そのためには特定の分野に縛られず、広い知識を収集し、広い視野で物を見る必要があると考える。

私は外国語学部出身でありながら、言語の研究の広がり認識して以来、独学で数学や情報科学についての知識を取り入れ、現在では情報科学の大学院で研究を行っている。心理学、神経科学的な言語に対する研究はまだ知識不足であるが、今後はそのような知識も取り入れる予定である。言語研究のそれぞれの分野の知見を取り入れ、またそれぞれに配信できるような言語研究の中心となるような研究者になりたいと考える。

【自己の長所】

上述のように言語に興味を持ち始めてから、いろいろな言語を独学で勉強した。特に興味があったアラビア語については2年間の独学の末、その努力が認められクウェート政府奨学金留学生に選ばれ、1年間クウェート国で衣食住すべて無料という好環境の中、各国の友達とアラビア語、アラビア文化の理解を深める機会に恵まれた。また、学部時代は言語以外にも歴史、数学、情報科学など多くのことに興味を持ちそれらについても独学した。このように私は学問に対して独学で学ぶということが習慣となっており、高い集中力で継続的に物事に専念できる人間であると考えている。

クウェート留学に応募した一義的な理由は、英語以外の新しい外国語の環境に自身をおいて、自身の脳がどのようにその言語の複雑な特徴（名詞の性など）を習得していくかを観察し、子供の言語学習とどのように違うのか比較したいと考えたことにあった。また学部時代の卒業論文は、情報科学、機械学習、自然言語処理の技術を専門のトルコ語に適用すればどうなるかということに興味を持ち、トルコ語意味解析の研究を、データ作成、プログラムの作成、論文執筆まですべて一人でやった。実験用のデータは、実際に複数のトルコ人にアンケートを取ることで作成した。これらの例で示されるように、私は好奇心が強く実験を行うことが好きな人間であると考えている。

②【自己評価する上で、特に重要と思われる事項】

私は、言語についての知識は誰にも負けないと自負している。 学部での自分の専門はトルコ語であり、それに加え留学で磨きかけたアラビア語とペルシア語の知識を活かし、難解なオスマン語（古典トルコ語）の読解において特に高い能力を示した。また上述のクウェート留学では、独学にもかかわらず、最初からアラビア語を専門とする各国の留学生に混じってアラビア語上級の授業を受け、無事に卒業証明書を得た。その他にもヨーロッパ、カフカス地方、中央アジアなどの多くの言語に取り組んでいたのも、留学中にはそれらを披露することで各国の友達と打ち解けるのにとっても役立った。英語についても高い能力を持ち945のTOEICスコアを保持している。日本語教育にも興味があり、独学して得た日本語学、言語学の知識を用いて日本語教育能力検定試験に受験、合格した。他にも学部時代は述語論理に基づいて作られた人工言語 lojban を学習するサークルに所属した。

博士前期課程では、自然言語処理の研究に注力した。その結果、採択率20%台の国際会議に2本の論文を通すことができた。 特にそのうちの一つの研究は国内でも高い評価を受け、第23回自然言語処理学会において優秀賞を受賞した。

申請者登録名 吉川将司