

2. 現在までの研究状況 (図表を含めてもよいので、わかりやすく記述すること。様式の改変・追加は不可(以下同様))

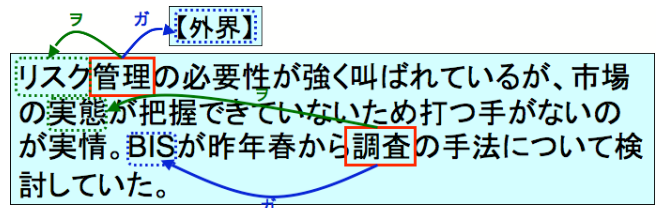
- ①これまでの研究の背景、問題点、解決策、研究目的、研究方法、特色と独創的な点について当該分野の重要文献を挙げつつ記述すること。
- ②申請者のこれまでの研究経過及び得られた結果について、問題点を含め①で記載したことと関連づけて説明すること。
 なお、これまでの研究結果を論文あるいは学会等で発表している場合には、申請者が担当した部分を明らかにしつつ、それらの内容を記述すること。

近年計算機の処理能力の向上と記憶容量の増大に伴い、膨大な量のデータに基づいた統計的な自然言語処理が盛んになった。特に形態素解析・構文解析といった領域では計算機を用いた統計的な手法が成功を収めており、成熟しつつある統計的なレベルの解析から、より深い自然言語理解のために、意味的なレベルの解析への研究が盛り上がりを見せている。高精度な意味解析は、自動要約や情報抽出、質問応答システムや言い換えといったアプリケーションを高度化するために必須の要素技術である。そういった技術の一つに、動作や状態を表す述語がそれぞれどのような要素(項)を伴うかを解析する述語項構造解析がある。我々は Gildea and Jurafsky (2002) の述語項構造解析に発想を得て、動作性名詞の項構造解析を行ってきた。

動作性名詞とはサ変名詞と動詞由来の名詞のことで、より深い自然言語理解のためには、動詞や形容詞だけでなく、動作性名詞に対しても項構造解析を行う必要がある。これらの名詞は事態を指すとき項構造を持つ。たとえば右図のような記事に対して

- 管理(する) [ガ:<外界>, ヲ:リスク]
- 調査(する) [ガ:実態, ヲ:BI S]

といった関係を解析の対象にする。



事態性とは文脈中で名詞がコト(動作)を指すかモノ(物体)を指すかといった意味的な違いに対応するもので、動作性名詞の項構造解析とは、名詞に事態性があるときに項構造を決定して項を同定する解析のことを言う。文脈に応じて動作性名詞に事態性があるかどうか判定する処理を事態性判別、項構造を決定して項を同定する処理のことを項同定と呼ぶ。

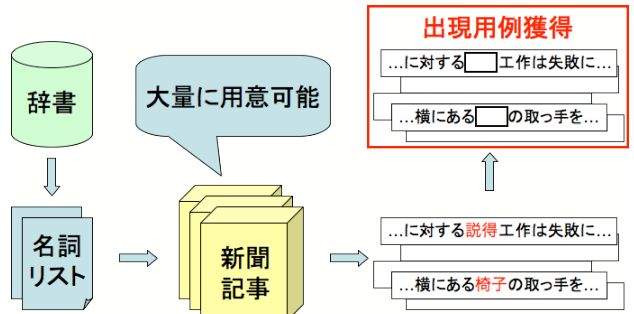
我々は動作性名詞の項構造解析のために、動作性名詞の項構造解析用のコーパス作成と、作成されたコーパスを用いた動作性名詞の項構造解析に取り組んできた。このコーパスは新聞記事(京大コーパス)を対象に動作性名詞について事態性の有無を判別し、事態性がある場合には項構造(必須要素となる表層格のガ格・ヲ格・ニ格)を手で付与したもので、申請者は2名の作業者と共にコーパスの仕様を策定し、現在までに780記事(6,500文)に対してタグ付けを行った。うち140記事は同一の記事に対し2名の作業者と独立にタグ付与し、タグの一致率を見たところ、事態性の判別に関してはほぼ両者で一致していたものの、ヲ格やニ格といった項の同定に関しては不一致が目立ち、項構造の選択に曖昧性があり、首尾一貫した項構造を選択することが比較的難しいことを示している。

事態性のタグは高い一致率で振れることが分かったので、我々は動作性名詞の項構造解析を

1. 事態性判別
2. 項構造解析(項同定)

という手順で問題を解くことにした。事態性判別は文中に出現する動作性名詞を事態性ありと事態性なしの2クラスに分類する問題なので、大規模な用例から事態性に関して曖昧性のない事例を用いて教師なしの学習を行うことができる。そこで、我々は事態性判別問題と項同定問題を分けて扱い、教師なし学習を行う有用性を示した。

教師なし学習においては、常に事態性のある用例しか持たない動作性名詞のリストと常に事態性のない用例しか持たない一般名詞のリストを用意し、新聞記事約1ヶ月分を使用してそれぞれの名詞の用例が含まれる文を取得し、事態性ありの事例を正例・事態性なしの事例を負例として学習する(右図)。学習には BACT (Kudo and Matsumoto, 2004) を用いた。BACT は文の構造を素性として用いることによって訓練事例の判別に効果が高い構造を判別ルールとして学習できる。



こうして事態性判別に有効な文構造を獲得して SVM (Vapnik 1998) を用いて事態性判別実験を行ったところ、文構造を用いなかった場合は精度 72.3%・再現率 58.7%、文構造を素性に用いた場合は精度 73.3%・再現率 80.2%となり、再現率が大幅に向上して事態性判別に大規模な用例から教師なしに獲得した文構造を用いることの有効性を示すことができた。

申請者氏名 小町守

この手法による事態性判別の問題点としては、項が文外に存在するために事態性なしと判定されたり、動作性名詞の周辺の文構造が事態性なしの事例に近いために事態性なしと判断されたりする事例があった。前者に関しては照応性の判定も含めて文外の項の同定も含めた事態性判別問題を考えることによって改善されると考えられるが、後者に関しては文の構造以外にも利用した事態性判別方法を考える必要があり、どのような手法で行うか検討中である。

関連する研究としては、NomBank (Meyers et al, 2004) はこういった名詞に関するコーパス作成を行うプロジェクトであり、動作性名詞の項構造解析の必要性も認知されるようになってきた。ただし、彼らは PropBank (Kingsbury et al, 2002) に従って文内の項のみを解析対象にしているのだが、特に日本語の場合は照応詞の省略(ゼロ代名詞)が頻繁に起こるため、文外の名詞句も項候補として扱う必要があると考え、我々はコーパスの仕様も文外の項に関する情報を付与するものにしていく。

日本語に関する名詞句の関係解析としては笹野ら(2005)・河原ら(2004)がある。彼らは事態関係だけでなく場所や属性・役割関係も含めたコーパスを作成して解析の対象としているが、我々は動作性名詞に焦点を当て、動作性名詞に特化した出現パターンを用いた解析を行っている点が特色である。動作性名詞に関係を絞ることで、大規模なコーパスから教師なし学習によって事態性判別に有効な出現パターンを獲得する手法を示した点が我々の研究の成果である。彼らの解析においては名詞句間に複数の関係の可能性がある場合の曖昧性を解消する問題を扱っていないため、たとえば事態性のない用例で使われている動作性名詞においても項構造があると判定され、項を探しに行くという問題点がある。

現在分かっている問題点は、名詞に事態性のあるとき正しく項構造が同定できるか複数の作業者間での一致率を見たところ、事態性の判別はほぼ人手による判断で行えるものの、事態性のあるときの格フレームの選択は作業者間で一致しないことが比較的多く、格フレームに関するタグの付与基準の見直しが必要である。格フレームに関する辞書を構築すること、そして格フレーム辞書を用いることで一貫性のある方法で項構造の選択と項の同定が行えるようなアノテーションツールの作成が必要であると考えられる。また、いまのところガ格・ヲ格・ニ格といった表層格を用いたコーパスの作成と項構造解析を行っているが、より深い意味解析を行うためには意味役割や語彙概念構造に基づく辞書やコーパスを用いた解析を行う必要があると考えている。

これまでの研究経過は 2006 年 3 月の言語処理学会第 12 回年次大会において口頭で発表した。

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つつ、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つつ記入すること。

動作性名詞の項構造解析は、文章中に表れる名詞間の関係のうち、明示的に事態を表す関係を対象とした解析を行う(研究業績(4))が、名詞間の関係は事態関係以外にも属性関係・役割関係・時間関係・場所関係といったさまざまな関係があり、これらを解析することも自然言語理解には必要なことである。また、役割関係など明示的には事態関係を表現していなくても総称的な事態が含まれているものがあり、これらを明らかにすることも重要であると考えている。

名詞間の関係解析は 90 年代に行われた MUC やそのあとに開催された ACE、そして GALE といったプロジェクトによって行われてきたが、これらは分野を限った関係解析に重きを置いており、コーパスの仕様が特定の関係に偏っているという問題がある。さまざまなアプリケーションに組み込める汎用性の高い要素技術としての意味解析を行うためには、分野に依存しないコーパスの設計と作成および解析手法の考案、もしくは分野に関するオントロジーなどの知識を比較的低コストで構築できる手法を開発する必要がある。

そこで我々はまず前者のアプローチを取ることにして、一般的なドメインのコーパスの仕様を作り、そのコーパスにおいて頑健な関係解析を達成することを第一に考えている。特に事態を表す関係においては、それぞれ事態に対応する項構造がどの語義に対応するのか、そしてその項構造の中でそれぞれの項はどのような役割を果たしているのか、自動で推定することを目指す。

(2) 研究目的・内容 (図表を含めてもよいので、わかりやすく記述すること)

- ①研究目的、研究方法、研究内容について記述すること。
- 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、具体的に記入すること。
- ③なお共同研究の場合には、申請者が担当する部分を明らかにすること。
- 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すること。

前節で述べたように、自然言語理解に当たっては名詞の項構造解析の研究が必須である。名詞の項構造解析に対するアプローチとして以下の2つの処理を提案する。

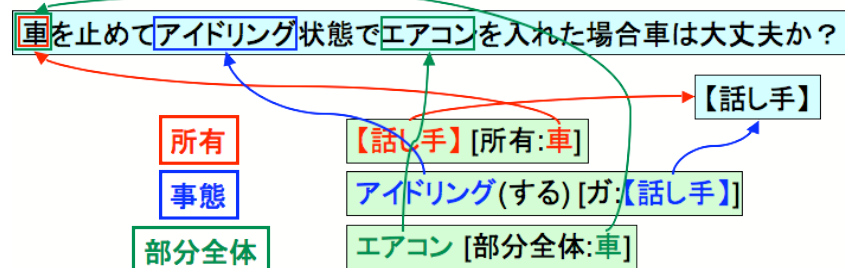
1. 事態の表す項構造推定

これまでの研究により、動作性名詞の事態性の推定は大規模な共起用例と文の構造情報を用いることによってある程度高い精度で行うことができることを示した。しかしながら、これまでに行った項構造解析では事態性のある場合の項構造は与えられているものとして、項同定を行う実験を行ったが、一般に事態性を持つ動作性名詞だけでなく、述語の用例がどの項構造で使われているのか推定することも未解決のサブタスクである。

そこで、我々は事態性判定問題と同様、項構造に曖昧性がある場合の曖昧性解消問題に取り組むだけでなく、動詞や動作性名詞が表す各事態に対してその項構造がどんな事態に対応しているのか半教師あり学習によって自動でラベルをつける手法も提案する。また、現在我々は表層格を用いたコーパスを作成しているが、動詞の格の交替現象などシステムの表層格と意味役割との対応付けの分析も含め、それぞれの項がどの意味役割を実現しているのか自動で推定する手法を提案する。PropBank や NomBank も各語義の項構造に対する項ラベルの一意性は保証されるものの、意味役割との対応付けは未解決の問題として残されているので、それぞれの語義に対応する項構造の内部に表れる項がどの意味役割を実現しているのか自動でラベルを付与する手法を考案する。

2. 名詞句の関係解析

名詞句の項構造には、事態性のある項構造だけではなく、さまざまな項構造が存在する。たとえば、以下のような文を考える。



この文が表している名詞間には、事態性のある動作性名詞とその項の関係だけではなく、所有者と所有物の関係、部分全体関係といったさまざまな関係が含まれている。こういった関係に加え、場所関係・時間関係など、任意性が高い項構造の解析も、意味解析技術を用いたアプリケーションを考えると重要である。名詞が項構造を持つのは典型的には事態性のあるときだが、このように明示的な事態性がなくても項構造を持つと考えてよい場合がある。このような名詞句間の関係は「A の B」という形の名詞句の用例に着目して大規模に獲得して関係性を推定するという方法が一般的だが、「A の B」という用例には関係の曖昧性があり、曖昧性がない事例を適切に選択して学習しないと使わないとむしろノイズにもなりうる。そこで我々は名詞句の関係解析においても関係性判定に有効な文構造を大規模な用例から推定し、必要なデータを能動的に選択して関係性判定に用いる手法を提案する。

また、研究計画の期間中には、名詞句の項構造解析とは別個に、NTT との共同研究に従事する予定である。この共同研究では日英・中英の句ベースの統計翻訳を行う予定であり、申請者はそのうちアライメントをつける句の切り出し方の研究に取り組む予定であるが、最終的には名詞の項構造解析を機械翻訳に応用する可能性も含めて研究の展開を検討している。

(3) 研究の特色・独創的な点

次の項目について記載すること。

- 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

FrameNet (Ruppenhofer et al, 2005) などを用いた述語項構造解析に関する先行研究では、述語に対応する項構造がどのような事態を表しているかは用例に従って人手で付与している。我々の研究計画にあるような、動詞も形容詞も名詞も統一的な枠組みで解析し、それぞれに対応する項構造がどういった事態を表しているか自動で推定する試みはまだ行われていない。また、PropBank (Palmer et al, 2003) では項構造の辞書を用意してコーパスのアノテーションを行っているが、それぞれの項がどの意味役割に対応するか推定するという取り組みは行われていない。

動作性名詞の項構造解析用のコーパスには英語の新聞記事コーパスである Penn TreeBank に対してタグ付けを行った NomBank (Meyers et al, 2004) があるが、これは文内の項に限定して情報を付与したコーパスであり、文外に項がある場合を考慮していない。我々は項の出現場所が文外になっている場合も解析対象にしており、特に日本語を解析対象にした場合に正しくモデルを構築できることが期待される。

一方、日本語の名詞句の関係解析としては笹野ら(2005)による名詞の格フレーム辞書を用いた関係解析がある。これは名詞句間のさまざまな関係を対象にしているという点では我々の研究と同じだが、関係性の判定問題を扱っていない。項構造の選択(取り得る項構造に多義性がある場合の多義性解消)を明示的に行う点と、項構造に対応する関係に自動でラベルをつける点が本研究の特色である。

本研究が完成すれば、現在自然言語処理以外の方野でも広く使われるようになった形態素解析や構文解析技術同様、意味解析も成熟した技術として用いることができ、文を超えた談話解析の研究が盛んになることが予想される。また、計算機を用いた意味解析で成功を収めることによって、言語学における意味論に解析可能性という新たな評価基準を与え、解析に応用できる意味論の研究を促進すると考えられる。

(4) 年次計画

(1年目)

表層格を用いたコーパスを完成させ、事態性を表す項構造の曖昧性解消の解析と項の同定のタスクを行う。現在仕様を計画して予備的にタグ付与中の名詞の関係解析用コーパスを用いた解析に着手する。文の構造を用いた名詞句の関係解析手法を提案する。名詞の項構造に関する辞書の作成も行い、辞書を効果的に用いたコーパスアノテーションツールの開発を行う。また、項構造辞書の作成を半自動で効率的に行う手法の研究を行う。

(2年目)

名詞の項構造解析を包括的に行う。表層格に基づいたコーパス・辞書の作成から意味役割もしくは語彙概念構造に基づいたコーパス・項構造辞書の作成へと向かい、名詞の項構造に対応する事態の自動推定と項に対する意味役割のラベルの自動付与の研究を行う。複合名詞の意味解析や名詞の照応解析といったさまざまな要素技術と組み合わせで解析を行い、それぞれの解析ツールの作成と公開を行う。

(3年目)

意味役割もしくは語彙概念構造に基づくコーパスと辞書を用いた項構造解析を行う。項構造解析をより頑健に行う方法を考案し、意見情報抽出などの実アプリケーションに組み込んで評価する。そしてそれまでの研究を踏まえて博士論文をまとめる。

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がある場合は項目に区分して記載すること。申請者にアンダーラインを付すこと）

(1) 学術雑誌等（紀要・論文集等も含む）に発表した論文又は著書（査読の有無を区分して記載すること。査読のある場合、印刷済及び採録決定済のものに限り、査読中・投稿中のものは除く）

①著者（申請者を含む全員の氏名を、論文と同一の順番とする）、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すること。なお、著者の所属・職については脚注に記載すること。

②採録決定済のものについては、それを証明できるものをP.8の後に添付すること。

(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すること）

著者（申請者を含む全員の氏名を、論文等と同一の順番で記載すること）、題名、発表した学会名、論文等の番号、場所、月・年を記載すること。発表者に○印を付すこと。

(4) 国内学会・シンポジウム等における発表

(3)と同様に記載すること。

(5) 特許等（申請中、公開中、取得を明記すること。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よい。）

(1) 学術雑誌等に発表した論文

[査読あり]

○小町守*、台湾植民地期の言語政策—安藤正次と二語併用の台湾、科学技術史、日本科学技術史学会、第8号、pp.31-55、2005年

(4) 国内学会・シンポジウム等における発表

[口頭・査読なし]

○小町守・飯田龍*・乾健太郎**・松本裕治***、共起用例と名詞の出現パターンを用いた動作性名詞の項構造解析、言語処理学会第12回年次大会発表論文集、pp. 821-824、2006年3月

* 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—助教授

*** 奈良先端科学技術大学院大学情報科学研究科—教授

申請者氏名 小町守

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すること。

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）

[研究者の志望動機・目指す研究者像・自己の長所]

研究者を志望するのは、自ら問題を発見し、自ら問題を設計し、自ら問題の解決に取り組む姿勢に強く惹かれるからである。学部時代までは問題解決に必要な基礎知識の習得と、問題解決の方法を学ぶことに集中しており、特定の問題設定の中で最適な解法を求めることを楽しんできたが、あまり現実世界で重要な問題に取り組むこともなく、パズルを解くような感じのおもしろさであったが、大学院に入ってから、問題を解くための道具立てがそもそもない状況から、現実的に解きたい言語学上の問題を分析し、解析可能なレベルにまで落とし込んで問題を設計し、コーパスを作って解析を行うといった研究に従事する機会に触れ、問題を解く楽しさから問題を設計する楽しさが少し見えてきたところである。そこでこれからも研究者としてどこまで自分が通用し、現実の問題を解いてこの分野の発展に寄与し、それを意識するにせよしないにせよ使ってもらう人たちに喜んでもらえるか試したい、というのが研究者を志望する動機である。

自分が目指す研究者というのは、当該分野の研究に従事するだけではなく、隣接する分野の研究者や応用面での方向性を見据えた技術者とのつながりを重視し、中長期的な視野で人と技術のマネジメントを行うことができる研究者である。自ら調べ自ら考え自ら手を動かす自立型の研究者になるのと同時に、自分が使う他分野の技術に対してユーザの立場で意見をフィードバックしていくことと、自分が開発する要素技術に関してそれを利用する他分野の研究者・技術者から開発者の立場でサポートを受け、頑健で広く使われる手法の研究・技術の開発を目指している。

自分の長所は、人とのつながりを大事にする開放性、さまざまなことに興味を持つ好奇心、そして一つの物事に取り組むときの集中力である。

一つ目は人とのつながりを大事にする点である。他大学との合同研究会や、ソフトウェアの開発者の会議があれば努めて参加するようにしており、自分がいて居心地がよくなるようコミュニティの雰囲気を風通しよくしたり、新しい人を勧誘して一緒に楽しく過ごしたりできるのは大きな長所であると考えている。

二つ目の長所はさまざまなことに対する好奇心である。元々科学哲学から科学史、言語学、そして自然言語処理へと専門分野を移り変わってきたが、分野の枠にとらわれず貪欲に知識を吸収するという姿勢で幅広く学び、有機的につなげていく能力は優れたものだと思う。

三つ目はなにか物事を始めると短期間に集中して没頭し、一定の成果を収めることがある。計算機に触れたのは大学に入学してからのことだが、大学3年のときにはオープンソースソフトウェアの開発者として活動するほどのめり込んだり、学部の後期課程に入ってから取り組んだ科学史の研究で論文を書いたり、現在の自然言語処理に移ってからの業績は少ないものの、知識を蓄える助走期間としては、今後次々と研究・開発の成果を公開していくつもりである。

[自己評価する上で、特に重要と思われる事項]

自分は研究や仕事の成果をオープンにしてみんなで共有することを強く意識しており、Linux というオープンソースソフトウェアおよび Fink というソフトウェア(パッケージ)管理システムの開発者の一員として大学時代から活動している。ソフトウェア開発者としては2005年7月に独立行政法人情報処理推進機構(IPA)が主催する未踏ソフトウェア創造事業(未踏ユース)に採択され、共同開発者の一員として「ユーザ参加型のパッケージ管理・斡旋システム」の開発に従事した。

ユーザと開発者の交流という観点からは、オープンソースカンファレンスや関西オープンソースといった開発者とユーザが交流する場にも積極的に参加してセミナーを行ったり、定期的に商業誌に記事を寄稿したりする活動が認められ、2005年11月には奈良先端科学技術大学院大学情報科学研究科特待生に選ばれて「オープンソース開発者支援プロジェクト」というプロジェクトで開発活動に対する支援を受け、また周囲のオープンソース開発者・開発者予備軍に対してのサポートを行っている。

また、自己の長所にも挙げたとおり、自分はさまざまなことに興味を持つ性質があるので、異文化の中での生活を体験してみようと学部時代オーストラリアに1年間留学した。そのとき出会った友人とも定期的に連絡を取っており、長期休暇の際に訪ねて旧交を温めたり、オープンソースソフトウェアの海外の開発者が日本に来るときは観光案内したりホストしたりと、いろいろな人と出会えることを楽しんでいる。

申請者氏名 小町守