

平成 21 年 4 月 16 日現在

研究種目：基盤研究 (C)  
 研究期間：2007～2008  
 課題番号：19500122  
 研究課題名 (和文) 情報集約に必要な文脈と名詞の属性を考慮した動詞の多義性解消モデルの構築  
 研究課題名 (英文) Construction of verb sense disambiguation model on the basis of context and noun attribute toward information summarization  
 研究代表者 竹内 孔一 (TAKEUCHI KOUICHI)  
 岡山大学・大学院自然科学研究科・講師  
 研究者番号：80311174

## 研究成果の概要：

動詞の語義の曖昧性を解消するために必要な事例と語義の曖昧性を記述するための動詞と名詞の定義集合を構築した。さらに事例をもとに動詞の語義を判別する自動付与システムの構築を行った。事例は新聞記事約1500文に対して約120語の動詞について動詞の語義、名詞の語義(日本語語彙大系)、名詞の意味役割の付与を行った(整理後公開予定)。さらに、事例を元に統計的学習モデルを利用して動詞の語義と名詞の意味役割を自動付与するシステムを構築した。

## 交付額

(金額単位：円)

	直接経費	間接経費	合計
2007 年度	2,400,000	720,000	3,120,000
2008 年度	500,000	150,000	650,000
年度			
年度			
年度			
総計	2,900,000	870,000	3,770,000

研究分野：計算機言語学・意味論

科研費の分科・細目：情報学・知能情報学

キーワード：知能情報処理, 自然言語処理, 語彙意味論, 意味役割, 語義付与コーパス

## 1. 研究開始当初の背景

大量のテキストから情報を集約するための基本技術の一つとしてある述語表現が他の述語表現と意味的な関係を捉えるシステムの開発が必要である。例えば「おにぎり/すしをにぎる」は「おにぎり/すしを作る」と言い換えてよいであろうし、「秘密/弱みを握る」は「秘密/弱みを知る」という意味とほぼ同じである

う。つまり、文脈(動詞と係り関係にある名詞)と動詞との組み合わせにより動詞の語義が異なり、さらにその語義に対して同様の概念を持つ動詞集合が存在する。しかしながらこうした処理を行うことができるシステムが一貫性を持って構築されてきていなかった。その大きな理由として動詞の語義ならびに名詞の語義を言い換えまで考慮して整理された辞書

ならびに事例データが今までに十分な量で存在しなかったことがあげられる。なぜならば膨大な名詞と動詞の組み合わせを人手で記述または整理することはほとんど不可能であるからである。よって動詞の語義とその語義における名詞の格のパターンが整理され十分な事例を持つ基礎データの構築が緊急的な課題であった。

## 2. 研究の目的

以上の背景から本研究では語義を付与したテキストデータを構築し、そのデータをもとに語義を自動付与するシステムを開発する。具体的には下記の3つを実現する。

### (1) 語義付与テキストデータの構築.

動詞の語義と係り関係にある名詞の語義を正しい係り受け解析結果に対して付与を行う。名詞の語義は①国語辞典による語義と②意味役割とよばれる動詞の語義からみた名詞の機能の2種類である。意味役割の定義は言語理論および付与事例における考察から構築し、事例付与を行う中で定義の精緻化を行う。

### (2) 語義付与をめぐる問題点の整理.

意味役割を含めた語義をテキストに付与する際、言語理論との接点からどのような問題があるのかを明確にする。さらに語義付与作業としての問題点を明らかにすることで、作成された語義付与テキストデータの利用可能性と価値を明らかにする。

### (3) 語義付与システムの開発.

入力テキストに対して語義の体系と事例をもとに動詞の語義と意味役割を自動で付与するシステムを開発する。開発したシステムの限界を明らかにすることで語義付与問題が現在の技術でどの程度困難であるのかを明らかにする。

## 3. 研究の方法

動詞の語義ならびに意味役割の基本集合とし

て本研究代表者が参加して構築した動詞項構造シソーラスを利用する。このシソーラスは公開されており、また基本語データベース(学研)の語義と対応関係があるため一貫した体系上での語義付与テキストデータを構築することができる。語義付与対象テキストとして新聞記事に対して人手で係り受け解析を修正されている京都大学コーパスを利用する。語義付与を行うテキストは選択した各動詞の語義に対して10例文ずつ集めるようにする。こうすることで、語義付与コーパスから動詞の語義の異なりがどのような名詞の異なりを得られるかを事例として残すことができる。ここで例文収集作業において新聞記事では使われない語義があり、幅広く収集するには対象テキストのジャンルを広くする必要があるのであることを明らかにした。

語義付与作業を安定して行うために語義付与補助ツールを開発する。作業者に対して他の付与事例や過去に付与した事例の検索が容易に行える環境を提供することで語義付与の揺れを防ぐことを目指す。シソーラスで定義されている意味役割集合は個別の動詞に直接関係する最小限の集合のみであるので、語義付与作業を通して意味役割の拡張を行う。同時に語義作業中に検出されるシソーラス上の語義の不足は修正を行う。つまり語義付与テキストデータ構築と同時にシソーラスの洗練を行うことで一貫した辞書と事例データの構築を行う。

自動語義付与システムは上記で構築した語義付与テキストデータを利用して統計的学習モデルを適用することで自動付与モデルを構築するこの際どのような特徴量を選択することが最適な精度を与えるかについて検討する。

## 4. 研究成果

本研究の成果は概略次の3つに集約できる。

(1) 動詞と名詞の語義を付与した意味役割付与データの開発.

京都大学コーパス約1500文に対し約120語の動詞について動詞の語義ならびに係り関係にある名詞の語義と意味役割を付与したテキストデータを開発した. 動詞の語義を中心に格パターンを記述する辞書の開発はIPALで行われていたが実際の記事に対して格パターンを記述したデータは作成されなかった. またEDR電子化辞書では意味役割付与テキストデータを構築したが語義毎ではなく記事毎に付与したため動詞の語義の曖昧性解消のために複数の事例を得ることは困難であった. 本語義付きテキストデータでは語義ごとに複数の例文を収集したため日本語の動詞語義を決定する文脈要素を調べるための基礎データを提供したことになる. 一方, 英語ではVerbNetやFrameNetのように大量のテキストデータに対して動詞の語義ならびに意味役割を定義して付与したデータが存在する. FrameNetを日本語で構築するプロジェクトが慶応大学を中心に始まっており, 今後新たな語義集合を付与したテキストコーパスが提供されることが予想される. こうした語義付与テキストデータは自然言語処理, 言語学分析の観点から必要なデータであり本研究成果はその先駆けと位置づけできる. 今後動詞の語義を増やしつつ事例を追加する予定である.

(2) 意味役割定義集合に対する理論的考察

意味役割集合は理論的には明確に定義されておらずいまだに決着を見ていない. この原因について歴史的経緯, 動詞分析における問題点, 意味役割に求められる処理の観点からの必要要件を整理し, 結局のところ人が事態を解釈する上で見方による異なりが必ず生じ, その結果意味役割だけでなく動詞表現そのものが個人の解釈により異なることが自然であ

ることを明確にした. この結果から意味役割は表現者の視点に沿って付与すべきであり, 元の事象の再構築は推論で行うべきであるという処理の枠組みを提案することができた. この考察結果は現段階では, シンポジウムや研究会での発表にとどまっているが, 今後論文として報告する予定である.

(3) 自動意味役割付与システムの開発.

(1) で開発した語義・意味役割付与テキストデータから事例を基に条件付確率場モデルを利用して意味役割を付与するシステムを開発した. システムを評価することで特徴量として動詞のフレームによる利用が効果的であること, 一方で名詞の語義と意味役割との相関が低いことを明らかにした. 現段階の意味役割精度は入力テキストの各要素に対して60%程度であるが, 意味役割付与システムの基盤が構築できた. 今後の精度向上を目指した開発を行う予定である. 現在規則ベースの意味役割付与システムを開発中であり, 安定した付与システムを構築できれば人手による語義付与テキストデータの開発を補助することが可能になると考えられる.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計0件)

〔学会発表〕(計2件)

1. 竹内孔一, 小山照夫. 動詞の語義と意味役割を付与したタグ付コーパスの作成, 電子情報通信学会, 言語理解とコミュニケーション研究会, 2008年11月10日, NICT麴町会議室.

2. 竹内孔一, 下村拓也. 動詞語義を推定するための語義付与コーパスの作成: 言語処理学会第14回年次大会, 2008年3月18日, 東京大学

〔図書〕(計0件)

[その他]

(1) 招待講演等

1. 竹内 孔一. 言語資源, 音声言語処理シンポジウム, 2008年12月09日早稲田大学小野記念講堂.
2. 竹内 孔一: 動詞と名詞の語義ならびに味役割を付与したグ付きコーパスの作成  
科研シンポジウム, 2008年9月26日熱海金城館.
3. 竹内 孔一, 黒田 航. 語彙意味論に基づく言語資源の構築. 言語処理学会第14回年次大会チュートリアル, 2008年3月17日, 東京大学.

(2) 解説記事等

1. 竹内孔一. 言語処理を指向した動詞項構造シソーラス, 月刊言語, Vol. 37, No. 8, pp. 58-64, 2008.
2. 竹内孔一: 言語情報処理における辞書と語彙概念構造語彙の意味と文法, くろしお出版, 由本陽子, 岸本 秀樹 (編), pp. 105-119, 2009.

(3) 関連報告書

1. 森本真衣子. 動詞項構造付与システムの構築, 岡山大学工学部情報工学科特別研究報告, 2009年2月9日.
2. 下村拓也. 動詞と名詞の概念体系を利用した意味役割モデルの構築, 岡山大学大学院博士前期課程論文, 2008年2月6日.

(4) ホームページ

動詞項構造シソーラス

<http://cl.cs.okayama-u.ac.jp/rsc/data>.

6. 研究組織

(1) 研究代表者

竹内 孔一 (TAKEUCHI KOUICHI)

岡山大学・大学院自然科学研究科・講師

研究者番号: 80311174