

## 科学研究費助成事業 研究成果報告書

平成 29 年 6 月 9 日現在

機関番号：15301

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26370485

研究課題名(和文) 言語処理及び言語分析を指向した大規模コーパスを利用した述語シソーラスの拡張

研究課題名(英文) Extending predicate-argument thesaurus with large-scale language resources for applying to natural language processing and linguistic analysis

研究代表者

竹内 孔一 (TAKEUCHI, Koichi)

岡山大学・自然科学研究科・講師

研究者番号：80311174

交付決定額(研究期間全体)：(直接経費) 2,400,000円

研究成果の概要(和文)：名詞項構造データの記述枠組を開発し名詞項構造データを構築した。項構造を番号で記述することで、付与された例文内の項を同定し、言い換え処理に必要な対応関係を記述する。また、名詞項構造データとして約5千例を作成し、そのうち1300例文(2,500箇所)について意味役割を付与した。品詞の異なる言い換えデータを構築し、述語項構造解析システムに取り入れ、文書に対して解析できるようにした。項構造解析システムに統計的学習モデルを導入し精度を向上させた。この内容を国際会議および情報処理学会論文誌で発表した。さらに含意認識タスクへの適用名詞の名詞項構造を利用した場合に認識精度が向上することを報告した。

研究成果の概要(英文)：We developed and published a new framework of nominal argument structure for Japanese. The key technique of describing arguments is number-based semantic roles, which enables us to identify each argument between paraphrased sentences in nominal predicates. We constructed 5,000 example sentences of nominal predicates; those of 13000 sentences have semantic role labels and links to the corresponding examples in the predicate thesaurus. We also constructed paraphrase data between different part-of-speech words and incorporate the paraphrase data into an argument structure analyzer. Besides, we improve the performance of the argument structure analyzer by applying a statistical learning method, and published the results in IPSJ Journal and the proceedings of International Conference PACLING 2015. We also show that the argument structure analyzer with nominal predicates enables us to improve performance of Japanese textual entailment task.

研究分野：自然言語処理

キーワード：項構造 名詞項構造 述語項構造 意味役割 述語項構造シソーラス

## 1. 研究開始当初の背景

言語処理において述語項構造は提案されているが日本語の名詞の項構造に関するデータの構築はIPAL辞書(1997)以降、構築されてこなかった。英語圏など複数の言語資源が構築されているが体系が異なっており相互リンクによりカバーしている状態である。例えば、単語間の意味的な関係を重視したWordNetには項構造が記述されていないがFramNetやNomBankには項構造が記載されている。しかしこれらは、WordNetの意味体系とは異なる記述枠組である。

一方、日本語では名詞の項構造に対する分析が言語学で近年展開され(西山2003,庵2007,影山2011)背景となる構造が提案されていた。また含意認識タスク名詞の項構造が必須であることから近年の理論と言語処理を考慮した項構造辞書の構築が求められていた。

## 2. 研究の目的

本研究代表者は言語処理における利用を目的とした述語項構造シソーラス(約1.1万語)を構築し、公開してきた。そこでシソーラス体系を拡張することによって一貫した記述枠組で名詞の項構造データ取り込むことを目標とする。この目標を中心として下記の5つの課題を設定した。

(1)名詞項構造の事例収集。言語学における分析を基に名詞の項構造例文を構築する。

(2)名詞項構造の記述枠組の開発。述語項構造シソーラスの体系、およびNomBank(Meyers他2004)など先行研究の体系、さらに理論からの提案(影山2011)を参考に、処理に利用できる付与可能な範囲の記述枠組を開発する。

(3)異なる品詞間での言い換え事例データを作成および収集する。

(4)項構造を自動で解析するシステムの精度向上を行う。述語項構造シソーラスの体系を基にした項構造解析システムを構築し、精度向上を行う。

(5)含意認識タスクに項構造を利用して認識

精度を向上させる。このために項構造を利用した含意認識処理システムを構築する。

## 3. 研究の方法

### (1)名詞の項構造例文の構築

当初、自動で作成された大規模格フレームデータからの半自動抽出を計画したがノイズ除去のコストが大きく、また例文が自動で取り出すことができなかったため、コーパスを基に名詞を収集し、作例により例文を構築する。

### (2)名詞項構造の記述枠組の開発

当初、述語項構造シソーラスと同じ意味役割ラベルを仮定していたが名詞の項のタイプと合致しない事例が出現した。そこで、NomBankを参考に項構造を番号で表す手法に切り替えた。シソーラスの意味役割との対応付けを行うことで一貫した体系を保ちつつ、記述可能な名詞の項構造の記述枠組を確立する。

### (3)シソーラスとの半自動マッチングツールによるデータ構築

名詞の項構造例文を項に分解し、対応する述語シソーラスの項目に関連づける半自動で対応づけるツールを構築する。これによりアノテーション作業を効率よく行う。

### (4)異なる品詞間での言い換え収集

当初、大規模テキストデータBCCWJに対して同時共起クラスタリングを利用した抽出を計画した。しかしながら実際適用すると抽出事例が少なく利用が難しいことが分かった。そこで類語辞書を参考に人手で異なる品詞間の言い換えデータを構築する。

### (5)処理システムの構築

項構造解析システムを構築して、処理精度の向上を目指しつつ、項構造解析システムを基にした含意認識システムを構築し、名詞を含めた項構造による解析を利用することで含意認識の精度が向上するかを確かめる。

## 4. 研究成果

(1)名詞の項構造データを記述する方法として、例文を作成し、項ごとに分解して意味役

割を番号で整理し、述語との言い換えがある場合には事例とともに項構造の対応も記述する枠組を提案し、第8回コーパス日本語学ワークショップで発表した。これにより、言語処理に利用可能でかつ構築可能な範囲の記述枠組を提示した。

## (2)名詞項構造データの構築

名詞項構造の例文として約5千例を作成し、そのうち1300例文(2,500箇所)について意味役割を付与し、述語との対応付けを行ったことを言語資源活用ワークショップで報告した。付与データの量は大規模では無いが、この構築作業を通して、記述可能で項構造解析システムと結びついたデータ構築方法を明らかにした。また、名詞項構造付与ツールを作成しているため、今後のデータ構築と容易になった。

品詞の異なる言い換えデータ(例えば「XとYは違いがある」と「XとYは異なる」)について人手により100表現を収集し、事例を含めて構造化し、述語項構造ソーラスに取り入れた(第7回コーパス日本語ワークショップで発表)。また述語項構造解析システムに取り入れ、文書に対して解析できるようにした。解析システムはホームページ上で公開した。

(3)項構造解析システムを構築し、統計的学習モデルを利用して精度を向上させた。先行研究では提案されてこなかった末尾表現を特徴量として取り入れることで精度が向上することを実験的に明らかにした。情報処理学会論文誌および国際会議PACLINGで発表した。

(4)名詞を含めた項構造の言語処理への適用可能性として物語理解への応用について論じた(下記の図書)。

## (5)含意認識タスクへの適用

名詞の項構造解析を述語項構造解析システムに実装し、含意認識タスクに適用した。含意認識における2つの例文に対して、係り受けベースのTreeKernelを利用して類似度を測定することで含意認識を行った。その結果、名詞

項構造を利用した場合に認識精度が向上することを明らかにし、言語理解とコミュニケーション研究会で発表した。

## 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3件)

石原靖弘, 竹内孔一, 係り元の末尾表現に着目した Hierarchical Tag Context Tree を利用した日本語意味役割付与システムの構築, 情報処理学会論文誌, 査読有, Vol.57 No.7, 2016, 1611-1626

Yasuhiro Ishihara, Koichi Takeuchi, Japanese Semantic Role Labeling with Hierarchical Tag Context Trees, K. Hashida and A. Purwarianti (Eds.) PACLING 2015, CCIS (Communications in Computer and Information Science) 593, 査読有, 2016, 21-32

DOI: 10.1007/978/981-10-0515-2\_2

Koichi Takeuchi, Masayuki Ueno and Nao Takeuchi. Annotating Semantic Role Information to Japanese Balanced Corpus, in Proceedings of MAPLEX 2015, 査読有, 2015

〔学会発表〕(計8件)

竹内孔一, 名詞項構造付与データの構築, 言語資源活用ワークショップ, 2017, P-D-2, 317-323

[http://pj.ninjal.ac.jp/corpus\\_center/lr/w/LRW2016\\_proceedings.pdf](http://pj.ninjal.ac.jp/corpus_center/lr/w/LRW2016_proceedings.pdf), 齋藤彰, 竹内孔一, コピュラ文を考慮した述語項構造解析器による含意認識, 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, 2017, 1-5

竹内孔一, BCCWJ への述語項構造ソーラスの付与による意味役割の検討, 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, 2015, 13-17

竹内孔一, オブジェクト指向に基づく言語インターフェース, ことば工学研究会, 2015

竹内孔一, 名詞の項構造データの構築, 第8回コーパス日本語学ワークショップ, 2015, 233-236

竹内孔一, 宮田周, 河村一希, 述語項構造を意識した名詞データの構築, 第7回コーパ

ス日本語ワークショップ予稿集，2015，  
143-146

竹内孔一，述語項構造シソーラスを意識  
した名詞の意味構造アノテーションのため  
の名詞意味構造の検討，第6回コーパスワ  
ークショップ予稿集，2014，51-56

竹内孔一，石原靖弘，竹内奈央，述語項構  
造シソーラスによる述語と名詞の構造化，  
人工知能学会全国大会，2014，215-0S-08b-1

〔図書〕(計1件)

Koichi Takeuchi et al., IGI Global,  
Computational and Cognitive  
Approaches to Narratology, 2016,  
191-212

〔産業財産権〕

出願状況(計0件)

名称：

発明者：

権利者：

種類：

番号：

出願年月日：

国内外の別：

取得状況(計0件)

名称：

発明者：

権利者：

種類：

番号：

出願年月日：

取得年月日：

国内外の別：

〔その他〕

竹内研のホームページ

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

6. 研究組織

(1)研究代表者

竹内 孔一 (TAKEUCHI, Koichi)

岡山大学大学院・自然科学研究科・講師

研究者番号: 80311174