

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 10 日現在

機関番号：12608

研究種目：若手研究(B)

研究期間：2013～2015

課題番号：25730131

研究課題名(和文) 同義・含意の関係にある格フレーム間の対応関係の自動獲得

研究課題名(英文) Automatic Acquisition of Correspondence Relation between Case Frames in
Synonym/Entailment Relation

研究代表者

笹野 遼平 (SASANO, RYOHEI)

東京工業大学・精密工学研究所・助教

研究者番号：70603918

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：本研究課題では、格フレーム間の対応関係の自動獲得を目的とし、主に受身および使役形の格フレームと能動形の格フレーム間の対応付け知識の獲得に取り組んだ。また、自動獲得した知識を受身文および使役文の能動文への変換における格変換タスクに適用し、受身文に対しては95.6%、使役文に対しては87.2%という先行研究の精度を有意に上回る精度で能動文における格を推定することが可能であることを示した。さらに、格フレームにおける各表層格の意味役割の分析を1つの目的として、格フレーム構築に使用したデータに基づく日本語二重目的語文の基本語順の分析に取り組み、基本語順は二格の意味役割により変化するなどの結論を得た。

研究成果の概要(英文)：This research project first developed a method for automatically acquiring knowledge about case alternations between the passive/causative and active voices, then confirmed the usefulness of the acquired knowledge through experiments of case alternations. In addition, this project conducted a corpus-based analysis of canonical word order of Japanese double object constructions to investigate the semantic role of each surface case of a case frame. The analysis suggested several conclusions including that the canonical word order varies depending on the semantic role of the dative argument.

研究分野：自然言語処理

キーワード：格フレーム 述語項構造 格交替 意味役割 基本語順

1. 研究開始当初の背景

近年、フレームを自動獲得しようという試みはいくつも行われており、日本語に対しても大量のテキストから表層格フレームを自動構築する手法が提案されている。しかし、構築された格フレームは異なる格フレーム間の項の役割の関係性は考慮していないため、含意関係認識など異なる動詞やヴォイス間の意味の対応関係性の認識が必要となるタスクに対してはそのままでは使えず、また、表層格のみ考慮しているため、意味役割の違いを認識することはできないという問題があった。

このため、これまでに能動形と受身形の格フレーム対を対象に、対応する格フレーム間の項の対応付け知識の自動獲得に取り組み、70億文にもおよび大規模テキストから得られる知識を用い、格フレームの用例間の類似度に着目することで、高精度な対応付けが可能であることを示した。しかし、テキスト中に出現する種々の同義関係や含意関係を認識するためには、能動形と受身形の格フレームの対応付けを行うだけでは十分であると言えないことから、さらに多くの格フレーム間の意味の関係性を明らかにすること、およびその実タスクにおける有用性を明らかにすることが必要であった。

2. 研究の目的

本研究課題では、(1)動詞・形容詞ごと、また、そのヴォイスや用法ごとに個別に構築された格フレーム間の対応付け知識を、格フレームの用例の類似度等に着目することにより自動獲得することで、異なる表層格構造をとる動詞・形容詞間の同義・含意関係や、共通の意味役割を持つ項を明らかにすることを目的とする。

また、(2)格フレーム構築に使用したデータに基づく日本語二重目的語文の基本語順の分析を行うことで、格フレームにおける各表層格の意味役割の手掛りを得ることも目的とする。

3. 研究の方法

(1) 個別に構築された格フレーム間の対応関係の自動獲得に関しては、主に受身および使役形の格フレームと能動形の格フレーム間の対応付け知識の獲得に取り組んだ。具体的には、受身形と能動形、および、使役形と能動形の間で格交替において交替する格はたかだか2つの格であり、格交替のパターンは限定的であることに着目し、起こりうる交替パターンを手手で記述した上で、約70億

文から構築した大規模格フレームから得られる類似度情報に基づいて格の対応付け知識の獲得を行った(図1)。また、自動獲得した知識を受身文および使役文の能動文への変換における格変換タスクへの適用することで、獲得した知識の有用性の検証を行った。具体的には、獲得した知識から得られる情報を機械学習に基づく手法の素性の1つとして追加することによる精度の変化を調査した。

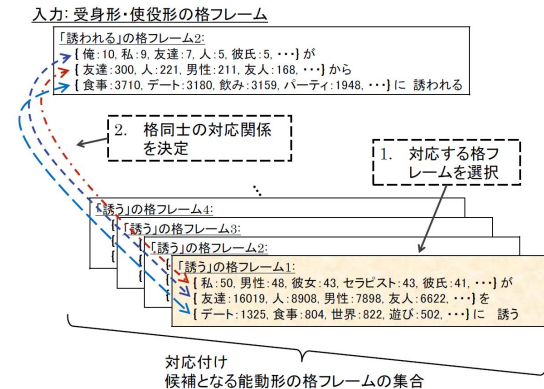


図1. 受身形・使役と能動形格フレームの対応付けの例

(2) 日本語二重目的語文の基本語順の分析に関しては、二重目的語構文の基本語順は実テキスト中における語順の出現割合(図2)と強く相関するという仮説に基づき、大規模コーパスを用いた日本語二重目的語構文の基本語順に関する考察を行った。ここで、実テキストを用いたアプローチにおいて問題となる点として、誤植や係り受けの曖昧性によって生じるノイズが挙げられるが、本研究では100億文以上から構成される大規模コーパスを用い、さらに確信度の高い部分のみを用いることで、信頼性の高い分析を可能とした。その上で、1)動詞によらず二重目的語構文の基本語順は「がにを」である、2)基本語順は動詞のタイプによって異なる、3)対象の動詞と高頻度に共起する格は動詞の近くに出現しやすい、4)基本語順は二格名詞の意味役割や有生性によって異なる、5)対象の動詞と高頻度に共起するヲ格名詞、二格名詞は動詞の近くに出現しやすい、という5つの仮説の検証に取り組んだ。

- 「がにを」語順優勢 コーパス中の出現割合
 - 言葉に愛情を感じる (97.5%)
 - タイヤに空気を入れる (99.1%)
- 「がをに」語順優勢
 - 女性をデートに誘う (99.6%)
 - 子供を風呂に入れる (98.6%)

図2. 大規模コーパスにおける語順の出現割合の例

4. 研究成果

(1) 約 70 億文から成るコーパスから構築された大規模格フレームから得られる類似度情報に基づき、受身および使役形の格フレームと能動形の格フレーム間の格の対応付け知識の獲得を行い、さらに、自動獲得した知識を受身文および使役文の能動文への変換における格変換タスクに適用することでその有用性を確認した。具体的には、新聞記事 2 万文を対象とした実験の結果、受身文に対しては 95.6%、使役文に対しては 87.2%(表 1) という先行研究の精度を有意に上回る精度で能動文における格を推定することが可能であることを示した。また、実験を通し、自動獲得された知識が使用できる場合は、人手で作成された種々の知識が不要となることを確認した。

モデル	受身文から 能動文	使役文から 能動文
最頻変換	0.883 (3159/3576)	0.781 (409/524)
提案手法 (学習なし)	0.938 (3353/3576)	0.878 (460/524)
先行研究 (学習あり)	0.944 (3376/3576)	0.836 (438/524)
提案手法 (学習あり)	0.956 (3417/3576)	0.878 (460/524)

表 1. 格交替推定実験の結果

また、獲得した知識に基づく格変換実験の拡張として、「問題を解く」と「問題が解ける」のような可能動詞化や、「ドアが閉まる」と「ドアを閉める」のような自他交替により表層格構造が変化した動詞を対象に格交替パターンの対応付けを行い、可能動詞化の場合は格助詞「を」から「が」への交替、自他交替の場合は格助詞「が」から「を」への交替を候補として考えることで、基本的に受身・使役と同様の手法を用いて行えることを確認した。

(2) 100 億文以上から構成される大規模コーパスに基づく日本語二重目的語文の基本語順の分析を行った結果、まず、動詞ごとの基本語順の分析結果として 1)日本語二重目的語文の基本語順は動詞により異なる、2)Pass タイプと Show タイプといった動詞のタイプは基本語順とほとんど関係しない、3)対象の動詞と高頻度に共起する格は動詞の近くに出現しやすい(図 3) という 3 つの結論を得た。

さらに、二格・ヲ格名詞の性質を考慮した基本語順の分析の結果、4)基本語順は二格の意味役割により変化する、5)対象の動詞と高頻度に共起するヲ格名詞・二格名詞は動詞の近くに出現しやすい(図 4)、という 2 つの結論を得た。

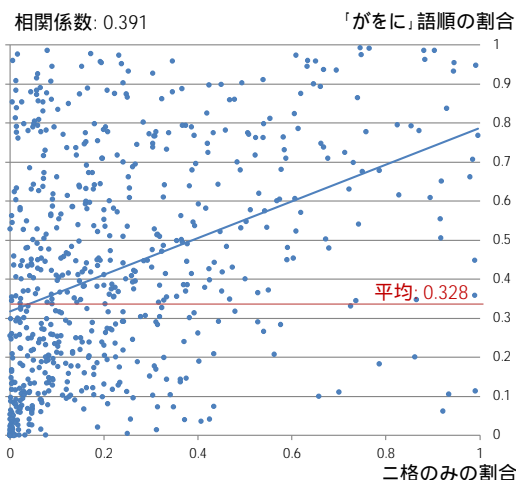


図 3. ヲ格と二格の一方のみ出現した用例に占める二格の出現率と、ヲ格と二格が両方出現した用例に占める「がをに」順の割合の関係

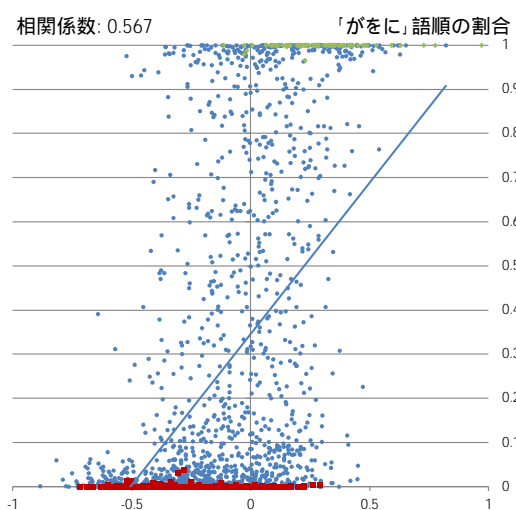


図 4. ヲ格名詞と動詞のNPMI から二格と動詞のNPMI を引いた値と、ヲ格と二格が両方出現した用例に占める「がをに」順の割合の関係

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 2 件)

1. 笹野遼平, 河原大輔, 黒橋禎夫, 奥村学: 受身・使役形と能動形間の格交替に関する語彙知識の自動獲得, 自然言語処理, Vol.21, No.6, pp.1207-1233 (2014.12), 査読有
2. 笹野遼平, 黒橋禎夫, 奥村学: 日本語形態素解析における未知語処理の一手法 -既知語から派生した表記と未知オノマトペの処理-, 自然言語処理, Vol.21, No.6, pp.1183-1205 (2014.12), 査読有

〔学会発表〕(計3件)

1. 笹野遼平, 奥村学: 大規模コーパスに基づく日本語二重目的語構文の基本語順に関する考察, 日本語学会 第151回大会, 名古屋(2015.11.28)
2. Ryohei Sasano, Daisuke Kawahara, Sadao Kurohashi and Manabu Okumura: Automatic Knowledge Acquisition for Case Alternation between the Passive and Active Voices in Japanese, In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP 2013), pp.1213-1223 Seattle, USA. (2013.10.19)
3. Ryohei Sasano, Sadao Kurohashi and Manabu Okumura: A Simple Approach to Unknown Word Processing in Japanese Morphological Analysis, In Proceedings of the 6th International Joint Conference on Natural Language Processing (IJCNLP 2013), pp.162-170, Nagoya (2013.10.15)

6. 研究組織

(1)研究代表者

笹野 遼平 (SASANO RYOHEI)
東京工業大学・精密工学研究所・助教
研究者番号: 70603918

(2)研究分担者

なし

(3)連携研究者

なし