

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 5月31日現在

機関番号：12608

研究種目：研究活動スタート支援

研究期間：2011～2012

課題番号：23800025

研究課題名（和文） 出現形と原形の格の対応付け知識を利用した原形格構造解析に関する研究

研究課題名（英文） Case Analysis for the Base Form with Knowledge of Case Alternation between the Surface and Base Forms

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

研究成果の概要（和文）：

本研究課題では、受身形で出現した用言の原形格構造を高精度に解析することを目的とし、同一の用言の対応する受身形と能動形の格の用例や分布の類似性に着目し、Web から自動獲得した大規模格フレームと、少数の受身形と能動形の格の交替規則を用いることで、受身形と能動形の表層格の対応付けに関する知識の自動獲得を行った。また、日本語構文・格解析システム KNP に機能追加する形で、省略・照応解析機能も含めた述語項構造解析器の公開を行った。

研究成果の概要（英文）：

This research project developed a method for automatically acquiring knowledge for case alternation between the passive and active voices in Japanese. By leveraging several linguistic constraints on alternation patterns and lexical case frames obtained from a large Web corpus, our method aligns an inputted case frame in the passive voice to a corresponding case frame in the active voice and finds an alignment between their cases. In addition, we released an anaphora and ellipsis resolution system as an extension of Japanese dependency and case structure analyzer KNP.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2011年度	1,300,000	390,000	1,690,000
2012年度	1,200,000	360,000	1,560,000
年度		0	0
年度			0
年度			0
総計	2,500,000	750,000	3,250,000

研究分野：工学

科研費の分科・細目：知能情報学

キーワード：自然言語処理, 述語項構造, 格構造解析, 格交替, 格フレーム

## 1. 研究開始当初の背景

テキスト中に出現した述語とその項の関係を認識する処理は述語項構造解析や格構造解析などと呼ばれ、計算機によるテキスト理解の実現のための重要な1ステップであると言える。しかし、一口に格構造解析と言っても、出力目的とする"格"としては、述語の出現形の格や、原形の格など複数考えられる。

出現形の述語項構造を扱う利点としては、先行する述語のガ格の格要素が後続する述語でもガ格の格要素になり易いなどといった構文的選好をそのまま利用できるため解析が比較的容易である点が挙げられる。

一方、原形の格を扱う利点としては、出現形に依らずに同内容であれば同じ格構造を出力されることから、テキストが表現する内容の同一性判定や、情報検索などに適している点が挙げられる。このため原形の格構造を求める研究は多く行われてきたが、構文的選好がそのままの形では利用できないなどの理由で、受身形などで出現した述語の格解析精度はそれ以外のものと比べ低い精度となっていた。

構文的選好を利用した原形格解析を実現する方法を考えると、出現形に対する述語項構造をまず求め、その後、原形格構造に変換するというアプローチが考えられる。一般に受身や使役形の格と原形の格の対応は用言ごと、用法ごとに異なるが、このような対応は大規模なコーパスから獲得可能であると考えられる。

## 2. 研究の目的

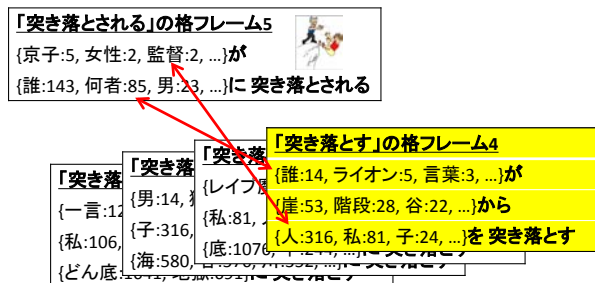
本研究の目的は (1) 受身形等、出現形の格構造と原形格構造を対応付ける知識の獲得、および、(2) 省略された格要素も対象とした出現形に対する格構造解析技術と、獲得した対応付け知識を組み合わせることで、これまで高い精度で解析することが難しかった受身形・使役形で出現した用言の原形格構造を高精度に解析するシステムを開発することである。

(3) また、開発した照応解析も含む述語項構造解析システムを広く一般に使用して貰うため、述語項構造・照応解析結果の閲覧性を考慮した上で、日本語構文・格解析システム KNP(<http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>)に機能追加する形で、システムの公開を行う。

## 3. 研究の方法

- (1) 受身形の格フレームと原形格フレームの対応付け知識の獲得に関しては、同一の用言の対応する受身形と能動形の格の用例や分布の類似性に着目し、Web から自動獲得した大規模格フレームと、人手で作成した少数の受身形と能動形の格の変換規則を用いることで、受身形と能動形の表層格の対応付けに関する知識を自動獲得する。格フレームとしては KNP に付属する形で公開されている Web69 億文から自動構築された格フレームを使用し、図 1 に示すように"ある受身形格フレーム"を入力とし、基本的に複数存在する能動形格フレームの中から"対応する能動形格フレーム"と"格の対応情報"を出力とする。

Input: ある受身形格フレーム



Output: 対応する能動形格フレーム+格の対応情報

図 1. 受身形格フレームと能動形格フレームの対応付け

具体的には受身形と能動形の間で格交替において交替する格は能動主体を表す格と、受身形でガ格となる格のたかだか 2 つの格であり、図 2~4 に示すように起こりうる交替のタイプはたかだか 20 種程度であるとの分析に基づき、起こりうる変換パターンを人手で記述した上で、69 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う。

### 1. [直接受身文] 対応する能動文でヲ格や二格で表される要素を主語とする受身文

- 大半が 推進派で 占められた  
大半を 推進派が 占めた
- 私が 先生から トークを 頼まれた  
先生が 私に トークを 頼んだ

- 能動主体を表す格: に, によって, から, で
- 受身形でガ格となる能動文の格: を, に

図 2. 直接受身文における格交替

2. [間接受身文] 能動文の表す事態に直接的に関わっていない人物を主語とする受身文

• 彼が 雨に 降られたのは...  
雨が 降った (NIL:彼)

- 能動主体を表す格: に
- 受身形でガ格となる能動文の格: NIL

図 3. 間接受身文における格交替

3. [持ち主の受身文] ヲ格やニ格で表されていた物の持ち主を主語とする受身文

• 彼が 男に 頭を バットで 殴られた  
男が 彼の 頭を バットで 殴った

- 能動主体を表す格: に
- 受身形でガ格となる能動文の格: の, NIL

図 4. 持ち主の受身文における格交替

- (2) 続いて、自動獲得した知識の有用性を確認するため、自動獲得した知識を受身文の能動文への変換における格変換タスクに適用し、自動獲得した知識が実タスクにおいて有用であることを確認する。
- (3) さらに、照応関係・述語項構造解析結果の閲覧性を考慮し、これらの解析結果の閲覧に適したオプションを追加した上で、日本語構文・格解析システム KNP に機能追加する形で、システムの公開を行う。

4. 研究成果

(1) 受身形の格フレームと原形格フレームの対応付け知識の獲得に関しては、格の対応付けの手掛かりとして、以下の3つの手掛かりを使用することにより、大規模かつ、高精度な格の対応付け知識を獲得することができた。

- ① 格用例の類似度: 対応する格の用例は類似しているという仮定に基づき、用例が類似している格に対応付けやすくする。
- ② 格用例数の分布の類似度: 対応する格の用例数の全体に占める割合は類似しているという仮定に基づき、格用例数の分布が似ている対応を選びやすくする。
- ③ 格変換の起こりやすさ: 格変換は格変換の種類ごとに起こりやすさが異なるという仮定に基づき、タグ付きコーパスから格変換の起こりやすさを算出し利用する。

(2) 自動獲得した受身形の格フレームと原形格フレームの対応付け知識を、格助詞変換タスクに適用し、自動獲得した知識の有用性の検証を行った。評価に使用するデータとしては、毎日新聞記事2万文中で受身形で出現した述語に係る格助詞を対象に、述語を能動形に変換した場合の格を付与した村田らのデータ (<https://alaginrc.nict.go.jp/case/index.html>) を使用し、2分割交差検定により評価を行った。

モデル	精度
[ベースライン]	0.8834 (3159/3576)
[村田モデル]	0.9444 (3377/3576)
[+自動獲得知識]	0.9564 (3420/3576)

表 1. 格変換の精度

表層格ごとにもっとも頻度の多い格に変換するベースライン、および、村田らのモデルとの比較結果を表1に示す。機械学習に基づく村田らの手法をベースに自動獲得した知識に基づく素性を新たに追加した結果、格変換の精度は0.9444から0.9564と有意に向上した。この結果から自動獲得した知識が実タスクにおいても有用であることが確認できた。

(3) さらに、日本語構文・格解析システム KNP に機能追加する形で、省略・照応解析機能も含めた述語項構造解析器の公開を行った。具体的には、共参照解析、ゼロ照応解析、連想照応解析機能を新たに KNP に追加し、さらに、照応解析結果の閲覧に適した出力フォーマットの追加を行った。

#	S-ID	SCORE	EID	解析結果
	1	-16.60015		太郎は 0 PERSON:太郎
				ケーキを 1 *
				買って 2 *ガ:太郎 [ヲ:ケーキ]
				食べていた。 3 *ヲ:ケーキ [ガ:太郎]
	2	-14.26750		
				とても 4
				おいそうだった。 5 *ガ:ケーキ [修飾:とても]

Saliency score ranking	
0:	太郎:1.500, 1:ケーキ:1.000
Predicate argument structure ranking	
0:	-32.625 【美味しい/おいしい:形1】 修飾:4;とてもガ:1;ケーキニ:×
1:	-33.813 【美味しい/おいしい:形1】 修飾:4;とてもガ:0;太郎ニ:×
2:	-34.116 【美味しい/おいしい:形4】 修飾:4;とてもガ:0;太郎ニ:×
3:	-34.219 【美味しい/おいしい:形10】 修飾:4;とてもガ:0;太郎ニ:×
4:	-34.384 【美味しい/おいしい:形5】 修飾:4;とてもガ:0;太郎ニ:×
5:	-34.614 【美味しい/おいしい:形1】 修飾:4;とてもガ:×

図 5. 述語項構造・照応解析結果の解析結果汎用表示ツールによる表示例

照応解析結果の閲覧に適した出力フォーマットとしては、照応関係・述語項構造に関連する解析結果のみ選択的に出力するオプション、および、図5に示すような解析結果汎用表示ツール (<http://nlp.ist.i.kyoto-u.ac.jp/nl-resource/taledisplay/index.cgi>) による表示を想定した出力オプションを追加し、照応関係・述語項構造解析結果を容易に閲覧できるようにした。

(2) 研究分担者  
なし

(3) 連携研究者  
なし

## 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1件)

- ① 笹野遼平, 黒橋禎夫: 大規模格フレームを用いた識別モデルに基づく日本語ゼロ照応解析, 情報処理学会論文誌, Vol. 52, No. 12, pp. 3328-3337 (2011. 12), 査読有

[学会発表] (計3件)

- ① 笹野遼平, 河原大輔, 黒橋禎夫, 奥村学: 構文・述語項構造解析システムKNPの解析の流れと特徴, 言語処理学会 第19回年次大会, 名古屋 (2013. 3. 13)
- ② 笹野遼平, 河原大輔, 黒橋禎夫, 奥村学: 大規模語彙的知識に基づく受身形と能動形の表層格の対応付け, 情報処理学会 第207回自然言語処理研究会, 札幌 (2012. 7. 27)
- ③ Ryohei Sasano and Sadao Kurohashi: A Discriminative Approach to Japanese Zero Anaphora Resolution with Large-scale Lexicalized Case Frames, In Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP 2011), pp. 758-766, Chiang Mai, Thailand, (2011. 11. 10).

[その他]

ホームページ等

URL: 研究成果を含むソフトウェア:

<http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>

## 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