B0TB2053

卒業論文

文間ゼロ照応解析のための格構造の類似度推定

大野雅之

20014年 3月 31日

東北大学
工学部 情報知能システム総合学科

文間ゼロ照応解析のための格構造の類似度推定*

大野雅之

内容梗概

言語を理解する上で,「誰が何をどうした」といった述語と項の関係の理解は非常に重要と言える．特に日本語では項の省略が頻繁に起きるため，省略を補完するゼロ照応解析に関する研究が盛んに行われているが，文をまたいだ照応関係を解析することは難しく，文脈情報を適切に利用することが求められている．本研究では，類似した項分布を持つ述語対を手がかりに文脈情報を扱う林部らの手法 [1] をベースに，格構造の類似度を推定する手法を提案する．

キーワード

自然言語処理，述語項構造解析，ゼロ照応解析

*東北大学 工学部 情報知能システム総合学科 卒業論文, B0TB2053, 20014 年 3 月 31 日.

目 次

1 序論

近年，Web などにより電子化された文書が広く使用され，それに伴い，情報抽出や機械翻訳などの自然言語処理の応用技術への需要が高まっている．これらの技術を実現するための基板となる照応・共参照解析や述語項構造解析に関して多くの研究がなされてきた．特に，日本語の文章に対する研究に着目すると，主語・目的語などの省略が頻繁に起こるという日本語の特徴から，省略を自動補完する**ゼロ照応解析**の研究が盛んに行われている．以下に省略が起きている文章の例を示す．

(1) 太郎$_i$は寝坊したので，(ϕ_i ガ）授業に遅刻した．

(1) に対して，省略されている「遅刻する」のガ格（**ゼロ代名詞** ϕ_i）の**先行詞**が「太郎」である，と解析することをゼロ照応解析という．特に，(1) のようにゼロ代名詞とその先行詞が同一文中に存在している事例を**文内ゼロ照応**と呼ぶ．文内ゼロ照応を解析する際には，統語的な情報を用いることで解析精度が向上することが，報告されている [2]．

一方，(2) の「供述する」のガ格のようにゼロ代名詞とその先行詞が異なる文に存在している事例を**文間ゼロ照応**という．

(2) 警察は太郎$_j$を窃盗容疑で逮捕した．付近の住民から多くの情報提供があり，早期逮捕につながったとみられる．「金に困っていた」と (ϕ_j ガ) 供述している．

文間ゼロ照応では，ゼロ代名詞が存在する文と先行詞が存在する文との間に談話が挿入されている場合が多く，統語的な情報を利用できない．また，前方に位置する全ての文が探索範囲になるため，先行詞候補の数が多くなる．これらの理由から，文内ゼロ照応解析と比べて、文間ゼロ照応解析は難しいとされている．そこで，本研究では文間ゼロ照応解析に焦点を当てる．

文間ゼロ照応解析では，先述の通り統語的な情報が利用できないため，文脈を捉えることが重要となっており，文章の顕現性首尾一貫性を用いた手法 [3] や，先行詞候補が以前の文中において述語の項となった回数を用いた手法 [4, 5] が存在

する．しかし，これらの手法では，(2) のように先行詞が主題として出現せず，また他の先行詞候補同様一度しか述語の項になっていない事例を正しく解析することはできない．林部らは類似した項分布を持つ述語対を手がかりに文脈情報を捉えるため，**格構造**を「助詞＋述語」と定義し項分布を比較することで，この事例に対しても解析を可能にした [1] が，述語の対象を動詞に限定していたため,「影響を与える」のように名詞側が意味をなす句に対して正しい項分布を捉えることができていなかった．本研究では，林部らが提案した格構造の類似度を用いる手法をベースに，意味が捉えられる範囲まで拡張した述語の項分布を扱うことで問題点を解消し，文間ゼロ照応の事例に対し先行詞同定の実験を行う．

以降，2 章では文間ゼロ照応解析に関する関連研究を取り上げ，本研究で使用する格構造の類似度について詳しく説明する．3 章では既存研究の問題点と改善策を提示する．4 章では実験設定と実験結果および考察について記述する．5 章では本研究で明らかになった点と今後の課題について記述する．

2　関連研究

文間ゼロ照応解析の既存研究の多くは機械学習を用いており，その素性として，文章の顕現性 [3] や，先行詞候補となる名詞句の情報 [4, 5]，述語間の関係 [6, 1] などが用いられている．本章ではこれらの既存手法について記述する．

2.1　顕現性に着目した手法

飯田らは，談話における話題の移り変わりを説明するセンタリング理論 [7] をもとに，Salience Reference List (SRL) を用いた手法を提案している [3]．SRL は，ハ格・ガ格・ヲ格・ニ格の項候補を１つずつスロットに保持するもので，以下の手順からなる．

1. 文章の先頭から，先行詞候補が各スロットに該当するか判別する．
2. 該当する場合，スロットに格納する．すでにスロットに項候補が格納されている場合，上書きして格納する．
3. 照応詞の述語の直前まで，以上の操作を繰り返す．

飯田らは SRL を
主題（ハ格）＞主語（ガ格）＞間接目的格（ニ格）＞直接目的格（ヲ格）＞その他
と順序付けることでガ格のゼロ照応解析に用いた．

2.2　先行詞候補の情報

文章中で一度述語の項になった名詞句は再び項になりやすいという知見がセンタリング理論 [7] の立場からも，統計的な観点 [8] からも得られているため，飯田らは先行詞候補が述語の項として使われた回数を機械学習の素性に用いた [4]．

今村らも同様の観点から，先行詞候補が項として使われたことがあるかどうかという真偽値を機械学習の素性として用いた [5]．

2.3 述語間の関係

2.2 節では先行詞候補単体で見た時に項になりやすいか否かといった情報を扱っていた．一方，述語から得られる情報を扱った手法も存在する．述語から得られる情報の代表的なものとして述語の選択選好が挙げられ，日本語語彙大系 [9] のような辞書資源が利用されている [10]．選択選好は,「逮捕する」のヲ格や「自首する」のガ格には「犯人」や「容疑者」のような名詞句が入りやすいといった，述語のどの格にどのような意味クラスの名詞句が入りやすいかという情報を示している．選択選好を用いることで，(3)a. の「自首する」の先行詞が「容疑者」であると解析できる．しかし，(3)b. の「自首する」の先行詞「花子」は一般名詞ではなく「容疑者」と同じ意味クラスではないため，選択選好では解析が行えない．

(3) a. 警察が容疑者を逮捕したと太郎は聞いた．逃亡に疲れ（ϕ ガ）自首したらしい．
b. 警察が花子を逮捕したと太郎は聞いた．逃亡に疲れ（ϕ ガ）自首したらしい．

そこで,「自首する」の前方に出現している「逮捕する」との関係のような述語間の関係を捉えることで解析を行う手法が提案されている [6, 1]．

2.3.1 項共有スコアの利用

飯田らは，スクリプト [11] に代表されるような事態の遷移とその遷移の中で共有される項の情報に着目し，述語対がどの程度項を共有しやすいかのスコアを算出した [6]．例えば，(4)a. では，述語「遅刻する」のガ格が省略されており，その先行詞は「太郎」である．ここで,「寝坊する」のガ格も「太郎」であるため，この文において「寝坊する」と「遅刻する」は項を共有している，といえる．一方,(4)b. では,「遅刻する」のガ格は「太郎」であるが,「怒る」のガ格は「次郎」となっており，この文において「遅刻する」と「怒る」は項を共有していない，といえる．

(4) a. 太郎は寝坊したので,（ϕ ガ）授業に遅刻した．

b. 太郎が遅刻したので，次郎は怒った．

飯田らは，述語項構造の関係タグが付与されたコーパスを利用し，係り受け関係にある述語対に対して，ガ格に共通の項を持つか否かを分類するモデルを作成した．この項共有分類モデルにより出力されるスコアを元に，述語対の項共有スコアを算出し，文間ゼロ照応解析の素性に加えた．

2.3.2 格構造の類似度の利用

林部らは,「格構造」を助詞と述語の組と定義し，2つの格構造が似たような項分布を持つとき,「格構造が類似している」と定義することで，格構造の類似度を用いてが述語間の関係を捉える手法を提案した [1].

(5) 警察は窃盗の容疑で太郎を逮捕した．最近，市内で被害が多発していた.「逃亡生活に疲れ，自首した」と供述している．

例えば，(5) の文章中の「自首する」のガ格は省略されており，先行詞は「太郎」である．ここで,「自首する」よりも前の文に出現している述語と項の組は「警察が逮捕する」,「太郎を逮捕する」,「被害が多発する」であり，これらの格構造の項分布を表1に示す．表1に示した項分布を見ると,「が自首する」と「を逮捕する」は似たような項を伴っており，これらの格構造は類似しているといえる．一方,「が自首する」と「が逮捕する」は項にとる名詞句の傾向が異なっており，格構造は類似していないといえる．このことから，自首した人が逮捕する可能性よりも，自首した人を逮捕する可能性が高いことが予想できるため，先行詞が太郎であると同定することができる．

表 1: 格構造の項分布（頻度順）

格構造	5715 が自首する		36560 が逮捕する		372219 を逮捕する		268351 が多発する	
項	791	＜人名＞	4404	警察	80366	＜数＞	36662	事故
	764	犯人	3574	県警	68876	容疑者	35170	事件
	490	男	2843	署員	54816	＜数＞人	14730	犯罪
	306	＜数＞人	1355	警視庁	23074	男	11530	トラブル
	288	＜数＞	1205	当局	7735	犯人	10926	被害
	274	容疑者	1200	警察官	6629	＜人名＞	6537	問題
	139	少年	1133	府警	5204	男性	6129	ケース

3　提案手法

本研究では，2.3.2 節で述べた手法をもとに，新たな格構造の類似度推定法を提案する．本章では，既存研究における問題点を述べ，その解決法を提案する．

3.1　格構造の類似度を利用する際の問題点

2.3.2 節で述べたように，格構造の類似度を利用する手法では,「助詞＋述語」を格構造と定義し項分布を比較することで類似性を測っていたが,「述語」の対象としていたものは「動詞」または「サ変動詞」（サ変接続の名詞＋する）のみであった．すなわち，(6) にある「影響を与える」といった「サ変名詞＋格助詞＋与える」に対しては，単に「与える」を述語として扱っていた．

(6)　(ϕ ガ) 影響を与える．

ここで,「影響を与える」の「与える」自体には内容的な意味はなく，その直前にある名詞「影響」が主な意味を持っている．そのため,「影響を与える」は「影響」を動詞化した「影響する」と同義になる．しかし，既存手法では (6) に対して,「が与える」の格構造を付与していた．表 2 から分かるように「が与える」の項分布は,「が影響する」と同様な項分布にならず，このような事例に対しては，本来捉えたかった項分布とは異なったものになっていた．

「影響を与える」のように,「実質的な意味を名詞に預けて，みずからはもっぱら文法的な機能をはたす動詞」を村木は「**機能動詞**」と名付け,「サ変名詞＋格助詞＋機能動詞」のように意味が捉えられる形にまとめたものを「**機能動詞結合**」と呼んだ [12]．他の機能動詞結合の例として「感銘を受ける」や「期待を抱く」などがある．

また，別の問題点として，曖昧性がある述語に付与された格構造の項分布が一様になってしまうという点がある．ここで，曖昧性がある述語とは，複数の用法が考えられる述語のことを意味している．例えば,「詰める」という述語は，(7)a,b のような用法がある．

表 2: 「が与える」と「が影響する」の項分布の比較（頻度順）

格構造	964006	が与える	510058	が影響する
項	20379	神	31508	それ
	20244	それ	12834	これ
	19030	神様	6828	環境
	16150	＜人名＞	6776	違い
	13686	成分	5873	変化
	9139	私	5789	要因
	8876	自分	5354	問題

(7) a. 店員が商品を箱に詰める.

b. 後続車が距離を詰める.

(7)a では「梱包する」と同様な意味をなしており，(7)b では「近寄る」と同様な意味として使われているため，これらの用法におけるガ格の項分布はそれぞれ「が梱包する」,「が近寄る」と類似したものになると考えられる.「が梱包する」と「が近寄る」の項分布は表 3 のようになり，これらは異なる名詞句を項にとることがわかる．既存手法では，(7)a,b の用法を区別せずに，どちらも「が詰める」の格構造を付与したため，項分布は表 3 のようになり,「が梱包する」と「が近寄る」の項分布を合わせたような分布になっている．このことから分かるように，曖昧性がある述語に対して 1 つの格構造だけ付与すると，一様な項分布になり項分布に特徴的な偏りが現れない.

表 3: 「が詰める」,「が梱包する」,「近寄る」の項分布の比較（頻度順）

格構造	45142	が詰める	2865	が梱包する	70515	が近寄る
項	5908	＜人名＞	297	スタッフ	4782	人
	1338	私	240	業者	3802	＜人名＞
	1254	＜数＞人	177	私	2457	男
	1209	人	104	お客様	2356	私
	997	選手	93	人	1227	男性

3.2 機能動詞結合の解消

本節では，3.1 節で述べた機能動詞結合の問題を解消するために行った処理について説明する．

機能動詞結合に関して，大竹 [13] や藤田ら [14] が，機能動詞結合を構成している「サ変名詞」を動詞化した形への変換を行った．以下に機能動詞結合の換言例を示す．

(8)a1. 監督が選手に指示を与えた．
a2. 監督が選手に指示した．
b1. 選手が監督に指示を受けた．
b2. 選手が監督に指示された．
c1. その映画は太郎に感動を与えた．
c2. その映画は太郎を感動させた．

(8)a では，機能動詞結合「指示を与える」が「指示する」に換言されているのに対し，(8)b では機能動詞結合「指示を受ける」が「指示される」と受動態に換言されており，機能動詞結合の換言は機能動詞によって態の変化が必要になる．(8)c を見ると，機能動詞は (8)a と同様に「与える」であるが，機能動詞結合を構成する名詞が (8)a では「指示」だが，(8)c では「感動」と異なっているため，換言後は「感動させる」と使役態になっている．このように，機能動詞結合を構成する名詞と動詞の組み合わせによって換言後の態を変換する必要がある．本研究では，大規模 web 文書から係り受け関係を抽出した項分布を用いるため，機能動詞を構成する名詞と動詞の組み合わせは膨大な種類になり，それらによる態の変換の規則を作成するのは困難である．

そこで，本研究では機能動詞結合の換言ではなく，機能動詞結合全体をひとつの述語とみなして格構造を付与するという手法をとった．具体例として，(8)a1 に対してガ格の格構造を付与することを考える．既存手法では「が与える」という格構造が付与されていたが，本手法では機能動詞結合である「指示を与える」をひとつの述語とみなし,「が指示を与える」という格構造を付与する．

既存手法で付与されていた「が与える」と本手法により付与された「が指示を

与える」，及びこの機能動詞結合と同義である「が指示する」の項分布を表 4 に示す.「が指示を与える」の項分布は「が指示する」の項分布と類似しており，本手法により機能動詞結合に対して本来の意味での項分布を持つ格構造を付与できていることが確認できる.

表 4: 機能動詞結合の解消前後での項分布の変化（頻度順）

格構造	964006	が与える	1361	が指示を与える	96708	が指示する
項	20379	神	143	＜人名＞	4245	＜人名＞
	20244	それ	120	監督	3907	私
	19030	神様	84	人間	3712	首相
	16150	＜人名＞	63	プレーヤー	2404	市長
	13686	成分	42	人	1790	先生
	9139	私	36	ユーザー	1558	監督
	8876	自分	34	コーチ	1490	医師

本手法を適用するためには，格構造を付与したい述語を機能動詞結合として扱うか否かを判定する必要がある．その判定を行うために，機能動詞結合になりうる「格助詞＋動詞」の対 223 個を文献 [12] に基づき選定することで機能動詞辞書を作成した．作成した機能動詞辞書を用いて，以下の条件を満たすものを機能動詞結合として扱った.

1. 述語のある文節とその直前の文節が係り受け関係にある.
2. 直前の文節が,「名詞句＋格助詞」の形である.
3. 「直前の文節の格助詞＋対象の述語」が機能動詞辞書内にある.
4. 直前の文節内の名詞句の主辞がサ変名詞である

機能動詞結合として扱うと判定された場合には，機能動詞結合全体をひとつの述語とみなして格構造を付与した.

3.3 曖昧性の解消

3.1 節で述べたように,「詰める」のように曖昧性がある述語に対して，どの用法に対しても「が詰める」という 1 つの格構造を付与すると項分布が広がってしまうという問題があった.

曖昧性のある述語に関して，再度「詰める」の例を以下に記す．

(9) a. 店員が商品を箱に詰める．

b. 後続車が距離を詰める．

(9)a,b において「詰める」は異なった意味で使われているが，これら両方の用法でガ格の他に，ヲ格またはニ格の項を保持していることが確認できる．このことから，曖昧性のある動詞に対してガ格以外の項（ヲ格またはニ格）を参照することで，どの用法で使われているか判別することができると考えられる．すなわち,「箱に」や「距離を」のようなニ格やヲ格の項を曖昧性のある述語とひとまとめにした状態でのガ格の項分布を見ることで用法を判断することができると考えられる．そこで，本研究では，ガ格以外に項を持つ述語に対して，ヲ格やニ格の項を埋めた状態での格構造を付与ることで曖昧性のある述語の曖昧性を解消する．本手法を適用すると，(9)b に対して新たに「が距離を詰める」という格構造が付与される．新たに付与された「が距離を詰める」と，本来の意味を表す「が近寄る」は，表 5 から分かるように似た分布をとっており，提案手法によって述語の意味を特定できたといえる．

ここで，(9)a に対して提案手法を適用すると,「が商品を詰める」と「が箱に詰める」という 2 つの格構造が新たに付与されるため，どちらの格構造を利用するか選択する必要がある．格構造の選択に関しては，次の節で詳しく述べる．

表 5: 「が距離を詰める」と「が近寄る」の項分布の比較（頻度順）

格構造	2865　が距離を詰める		70515　が近寄る	
項	390	＜人名＞	4782	人
	85	男	3802	＜人名＞
	56	選手	2457	男
	52	車	2356	私
	42	人	1227	男性

3.4　ゼロ照応解析への適用

本節では，格構造の類似度の算出法を示し，新たに付与した格構造のゼロ照応解析への利用方法について記述する．

3.4.1 格構造の類似度

2.3.2 で述べたように，格構造の類似とは格構造の項分布が類似していることと定義されており，項分布の類似度の算出法を定める必要がある．本研究では，2 つの項分布 p,q をベクトルとみなすことで，項分布の類似度 $Sim(p,q)$ をコサイン類似度を用いて以下のように定義する．

$$Sim(p,q) = cos(p,q) = \frac{\sum p(x)q(x)}{\sum p(x) \sum q(x)}$$

コサイン類似度は，2 つのベクトルがなす角度を考えることで類似度を扱っている．

3.4.2 先行詞同定

先行詞候補の中から先行詞を選出するための手掛かりとして，各先行詞候補に対して以下で定義する類似スコアを付与した．先行詞候補 n を項に持つ述語の格構造 h_i の集合を格構造履歴 $H = h_1, h_2, ..., h_n$ と定義し，以下の式で求めた値を先行詞候補 n と着目している述語の格構造 p との類似スコアとした．なお，n と照応関係にある名詞句の格構造履歴も H に含む．

$$Score_{sim}(p,H) = \max_i Sim(p,h_i)$$

例えば，名詞句「容疑者」が「逮捕する」のヲ格と「盗む」のガ格の項であった場合，格構造「が自首する」との類似スコアは

$$\begin{aligned}
&Score_{sim}(\text{が自首する}, \{\text{ を逮捕する}, \text{が盗む }\}) \\
&= max(\{Score_{sim}(\text{が自首する}, \text{を逮捕する}), Score_{sim}(\text{が自首する}, \text{が盗む})\}) \\
&= max(\{0.504472, 0.670937\}) \\
&= 0.670937
\end{aligned}$$

となる．

また，本研究では既存手法での問題点を解決するため，述語に対して新たな格構造を付与した．これらは，機能動詞結合への付与とが格以外に項を持つ述語への付与にわけることができ，類似度算出時のそれぞれの扱い方を説明する．

機能動詞結合は，構成している動詞自体は意味を持っていないという特徴があるため，既存手法による「格助詞＋動詞」で定義される格構造の分布を用いて類似度を算出することは好ましくない．従って，機能動詞結合に対する格構造の類似度算出では，提案手法によって付与した格構造を用いる．

ガ格以外に項を持つ述語に対しては，項をひとつ埋めた状態での格構造を付与したが，(9)a のように述語がヲ格とニ格両方の項を持つ場合,「が商品を詰める」と「が箱に詰める」といった 2 つの格構造が付与される．そこで，ガ格以外に項を持つ述語では，類似度算出を行う際，既存手法で付与された格構造と提案手法で付与された格構造全ての組み合わせに対して網羅的に格構造の類似度を算出し，その中で値が最大のものを格構造の類似度として選ぶ．

(10) 警察官が男を取り押さえ，少年は救出された．(ϕ ガ）刃物を持って，家に立て籠もってから 20 時間経っていた．

例えば (10) の文で，述語「持つ」はヲ格に「刃物」を項に持っているため先行詞「男」との類似度は $Score_{sim}$(が持つ, を取り押さえる) と $Score_{sim}$(が刃物を持つ, を取り押さえる) の 2 つが考えられるがそれぞれの類似度を計算すると

$$\begin{aligned} Score_{sim}(\text{が持つ, を取り押さえる}) &= 0.187310 \\ Score_{sim}(\text{が刃物を持つ, を取り押さえる}) &= 0.685678 \end{aligned}$$

となるため,「男」に付与される類似スコアは 0.685678 となる．同様に先行詞候補「警察官」と「少年」に付与される類似スコアはそれぞれ 0.276867,0.246791 となり類似スコアが最も大きい選ぶことで「男」を先行詞と解析できる．

4　評価実験

提案手法の元となる林部らの既存研究では，格構造の類似度を素性のひとつとして述語項構造解析器を作成していたが，本実験では機能動詞結合と曖昧性のある動詞によって生じる問題の解消を行うことの有効性を調査するために，林部らが提案した格構造の類似度のみを用いた場合での先行詞候補の順位付けを行うことで先行詞同定の精度を確認する．

4.1　実験設定

提案手法の有効性を調査するため，林部らが提案した格構造の類似度のみを用いたモデルをベースラインに設定し，それらに

- 3.2 節の手法で機能動詞結合の問題点を解消
- 機能動詞結合の問題点の解消と，3.3 節の手法で曖昧性へ対応

を行ったモデルの 3 つで先行詞候補の順位付けをし比較した．実験では，先行詞をどの程度上位に順位づけできたかを以下の式を用いて評価した．

$$MRR = \frac{1}{N} \sum_{n \in N} \frac{1}{rank(n)}$$

ここで，N は事例の数を表し，$rank(n)$ はある事例 n における先行詞の順位を表す．

格構造の項分布は，web から収集した約 60 億文に対して CaboCha 0.66[15] を用いて形態素解析・係り受け解析・固有表現解析を行ったものより，述語と名詞句の格助詞を介した係り受けの計 5,895,225,186 対を抽出しその頻度[1]を用いた．また，曖昧性のある述語に対して付与する格構造の項分布は，述語に対して格助詞「が」以外に格助詞を介して名詞句が係り受けがある計 169,260,929 事例の頻度を用いた．

評価には，NAIST テキストコーパス 1.5 [16] を対象に文間ゼロ照応の関係のうち頻出するガ格の 7854 事例を利用した．NAIST テキストコーパス 1.5 は京都

[1] ノイズ除去のため頻度 5 以上の対を利用した

テキストコーパス version 4.0[2] で利用されている 95 年 1 月 1 日から 17 日までの全記事（約 2 万文），1 月から 12 月までの社説記事（約 2 万文），計約 4 万文に対して，述語の格関係，事態性名詞の格関係，名詞間の照応関係をアノテートしたコーパスである．

4.2 結果

それぞれのモデルにおける MRR を，先行詞が 1 位に順位付けられた事例数と共に表 6 に示す．機能動詞結合に対して新たな格構造を付与したことで MRR が 0.001 と僅かに上昇し，同時に曖昧性への対処を適用することでベースラインから 0.006 上昇した．また，機能動詞結合への対処によって先行詞を 1 位に順位付けできた事例数はベースラインと比較すると 2 事例多いだけであったが，曖昧性への対処を適用することで 61 事例に増えた．

表 6: 文間ゼロ照応における MRR, 及び先行詞が 1 位に順位付けられた事例数

モデル	MRR	1 位に順位付けられた事例数
ベースライン	0.733	4688
機能動詞結合の解消	0.734	4690
機能動詞結合の解消＋曖昧性への対処	0.739	4749

4.3 考察

提案手法による，機能動詞結合，及び曖昧性への対処を行うことで，先行詞候補の順位付けにおける MRR は上昇したが，その上がり幅は僅かなものであった．先行詞を 1 位に順位付けできた事例数で比較すると，機能動詞結合への対処のみの場合は MRR 同様僅かに増加しただけであったが，曖昧性への対処を適用すると 1 位に順位付けできた事例数を増やせた．本実験では，格構造の類似度のみで先行詞候補の順位付けを行っており，先行詞が 1 位に順位付けられた事例は格構

[2]http://nlp.ist.i.kyoto-u.ac.jp/index.php

造の類似度のみで先行詞同定が行えたことを意味しているため，ベースラインよりも改善がみられたと言える．

提案手法を適用することによりベースラインと比較して順位がどう変動したかを事例単位で調査したものを表 7 に示す．提案手法を適用することで，全体の約 9 %の事例で順位の向上がみられた．しかし，約 5 %の事例で順位が悪化した．

表 7: 事例単位での順位変動

モデル	順位が向上下事例数	順位が悪化した事例数
機能動詞結合の解消	552	295
機能動詞結合の解消＋曖昧性への対処	730	447

順位が悪化した理由として，web から抽出した格構造の項分布が疎になったことがあげられる．提案手法では，ガ格以外に項を持つ述語に対して項を固定した格構造を付与する際に，単純に項と動詞を繋げたものを述語として扱っていた．すなわち,「が距離を詰める」と「が間合いを詰める」のように同じ用法であっても付与される格構造は異なっていた．その結果，項となる名詞句の種類が少なく，項分布が疎になってしまったため，類似度算出に悪影響を与えたと考えられる．この問題に対して，固定された項となる名詞句をクラスタリングするといった対処法が考えられる．クラスタリングすることで「距離」と「間合い」のように類似した名詞句をまとめあげることができ，同じ用法の場合には同一の格構造が付与されるようになり，項分布が疎になる問題を解消できると考えられる．

本実験では，格構造の類似度のみを用いて先行詞候補の順位付けを行ったが，以前の文中で一度も述語の項になっていない先行詞には類似スコアを付与できないため解析対象から外していた．よって本実験で順位付けが行えたのは文間ゼロ照応である 21151 事例中 7854 事例と 37 %程であった．本手法が適用できない事例に対しても解析を行うためには，従来の手法と同様に他の素性と組み合わせて解析を行う必要がある．

5　結論

本論文では，既存手法を元に問題点を解消した格構造の類似度を用いて文間ゼロ照応の事例に対し，順位付けによる先行詞同定を行った．機能動詞結合や曖昧性のある動詞に対しても適用できるよう格構造を拡張することで，僅かではあるが先行詞同定の性能を向上することができた．提案手法では項分布が疎になるといった新たな問題点が生じたため，今後の課題として格構造を付与する際の項と述語のまとめあげにおいて，項をクラスタリングすることが挙げられる．また，本研究では他の素性と組み合わせた評価を行っていないため，提案手法による格構造の類似度を述語項構造解析器の素性として加えることによる性能の変化を調査する必要がある．

謝辞

本研究を進めるにあたり，ご指導頂いた乾健太郎教授，岡崎直観准教授に感謝致します.

研究の進め方や本論文の作成など様々な場面で親切に指導して下さった研究員の井之上直也氏に感謝いたします.

日頃の議論を通じて多くの知識やご指摘を下さった乾・岡崎研究室の皆様に感謝します.

参考文献

[1] 林部祐太, 小町守, 松本裕治. 文脈情報と格構造の類似度を用いた日本語文間述語項構造解析. 情報処理学会研究報告. SLP, 音声言語情報処理, Vol. 2011, No. 10, pp. 1–8, may 2011.

[2] Ryu Iida, Kentaro Inui, and Yuji Matsumoto. Exploiting syntactic patterns as clues in zero-anaphora resolution. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, ACL-44, pp. 625–632, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics.

[3] Ryu Iida, Kentaro Inui, Hiroya Takamura, and Yuji Matsumoto. Incorporating contextual cues in trainable models for coreference resolution. In *In Proceedings of the EACL Workshop on The Computational Treatment of Anaphora*, pp. 23–30, 2003.

[4] 飯田龍, 乾健太郎, 松本裕治. 文脈的手がかりを考慮した機械学習による日本語ゼロ代名詞の先行詞同定 (自然言語). 情報処理学会論文誌, Vol. 45, No. 3, pp. 906–918, mar 2004.

[5] Kenji Imamura, Kuniko Saito, and Tomoko Izumi. Discriminative approach to predicate-argument structure analysis with zero-anaphora resolution.

[6] 飯田龍, 徳永健伸. 述語対の項共有情報を利用した文間ゼロ照応解析. 言語処理学会第 16 回年次大会 発表論文集, Vol. 16, pp. 804–807, mar 2010.

[7] Barbara J. Grosz, Scott Weinstein, and Aravind K. Joshi. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, Vol. 21, pp. 203–225, 1995.

[8] Niyu Ge, John Hale, and Eugene Charniak. A statistical approach to anaphora resolution. In *In Proceedings of the Sixth Workshop on Very Large Corpora*, pp. 161–170, 1998.

[9] 池原悟. 日本語語彙大系, 1999.9 1999.

[10] Ryu Iida, Kentaro Inui, and Yuji Matsumoto. Zero-anaphora resolution by learning rich syntactic pattern features. 2007.

[11] Roger C. Schank and Robert P. Abelson. *Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures*. L. Erlbaum, Hillsdale, NJ, 1977.

[12] 村木新次郎. 日本語動詞の諸相. ひつじ書房, 1991.

[13] 大竹清敬. 機能動詞結合の換言に伴う連体修飾表現の変換. 言語処理学会第 11 回年次大会 発表論文集, Vol. 11, pp. 337–340, mar 2005.

[14] 藤田篤, 降幡建太郎, 乾健太郎, 松本裕治. 語彙概念構造に基づく言い換え生成 : 機能動詞構文の言い換えを例題に (自然言語). 情報処理学会論文誌, Vol. 47, No. 6, pp. 1963–1975, jun 2006.

[15] 工藤拓, 松本裕治. チャンキングの段階適用による日本語係り受け解析. Vol. 43, No. 6, pp. 1834–1842, 2002.

[16] 飯田龍, 小町守, 井之上直也, 乾健太郎, 松本裕治. 述語項構造と照応関係のアノテーション: Naist テキストコーパス構築の経験から. 自然言語処理, Vol. 17, No. 2, pp. 25–50, 2010.