

視線情報を用いた日本語述語項構造解析モデルの パラメータ推定

牧 諒亮^{1,a)} 西川 仁^{2,b)} 徳永 健伸^{2,c)}

概要: 本稿では、日本語述語項構造解析モデルのパラメータ推定において視線情報を利用する方法を提案する。文章内の言語的な情報のみだけでなくアノテーション時の作業者の視線から得られる情報を取り入れることで解析の精度を向上させることが目的である。ある述語の項を探すアノテーション作業者が頻繁に見る候補は項らしい候補であると考え、各項候補に対する作業者の停留は項らしさを反映するという仮説を立てた。この仮説にもとづき、述語項構造解析モデルのパラメータ推定に作業者の視線を取り入れた。評価実験により、視線情報を取り入れることで述語項構造解析の精度を 0.05 ポイント向上する結果を得た。

1. はじめに

今日、教師あり機械学習は自然言語処理において重要な手法の1つとなっており、そのためにさまざまな情報を付与したアノテーションコーパスが構築されてきた [1]。コーパスに付与された情報は、与えられた課題に対してモデルを構築する際に、モデルの出力として期待されるものであり、それゆえ訓練データとして利用される。しかしながら現在の枠組みにおいては、アノテーションの結果として付与された情報のみが利用されるに留まっている。本稿ではアノテーションの結果付与された情報だけでなく、アノテーション過程におけるアノテーション作業者の振る舞いをモデルの構築に利用する手法を提案する [2]。

本研究では日本語における述語項構造解析を対象とする。日本語における述語項関係は多くの場合、格助詞によって示されており、それゆえ表層格を対象としてアノテーションを施すコーパスが構築されてきた [3], [4]。これらのコーパスでアノテーションの対象となる格は必須格とされるガ格・ヲ格・ニ格が中心であり、これらの格要素を予測する述語項構造解析器が構築されてきた [5], [6]。しかしながら、日本語の文章においては省略が珍しくないため、述語と同一文内に格要素が現れない場合もあり、その際には格要素を同定するために文章を遡る必要がある。この点が英語における述語項構造解析と異なる点である。さらに、受身や使役の助動詞によって格の交替が生じることがあり、アノテ

ション基準によっては、述語項構造解析の過程で本来の格に戻す必要がある。この点において、日本語述語項構造解析は英語における意味役割付与 [7] と類似している。最後に、日本語の名詞には、動詞や形容詞と同じように項を持つ名詞があることが指摘されており、それらは事態性名詞と呼ばれている [8]。事態性名詞についても述語項構造を同定する必要があるため、述語項構造解析の対象として扱われている。述語項構造が付与されたアノテーションコーパスとしては NAIST テキストコーパス (NTC) [3] や BCCWJ-DepParaPAS [4], [9], [10], [11], [12] が公開されている。本研究においてはコーパスとして BCCWJ-DepParaPAS を利用する。

本稿では述語項構造解析の性能を向上させるため、述語項関係をアノテーションする作業者の振る舞いとして視線情報を利用し、モデルの構築を行う手法を提案する。既存の述語項構造解析においては、与えられた述語に対して人間がアノテーションした項を正解の項として、機械学習によるモデル構築を行うという手法がとられてきた。言い換えれば、正解の項のみが正例として扱われ、それ以外の候補は負例として扱われてきたということである。アノテーション作業者の作業中の視線を観察してみると、最終的な判断をするまでに様々な候補を見ていることがわかる。最終的な判断において選ばれなかった候補は、その文章における対象述語の項ではないものの、頻繁に見ていた候補については他の文章において同じ述語の項となる可能性がある。提案手法のアイデアは、このような候補から得られる情報をモデルの訓練段階で利用する点にある。

次の例文を考えてみよう。

¹ 東京工業大学 大学院情報理工学研究所

² 東京工業大学 情報理工学院

a) maki.r.aa@m.titech.ac.jp

b) hitoshi@c.titech.ac.jp

c) take@c.titech.ac.jp

- (1) [私_が]は昨日、友達とランチを食べに行きました。
値段にも味にも満足しました。
- (2) 私は昨日、[友達_が]とランチを食べに行きました。
値段にも味にも満足してくれたようです。

これらの文章の組は、各文章の2文目の末尾が異なることを除いて同じ要素から構成されている。一方、下線を施した述語「満足する」のガ格の項を見てみると、(1)では「私」、(2)では「友達」であり異なっていることがわかる。これらの文章を利用して「満足する」に対する訓練事例を生成する際には、(1)における「私」と(2)における「友達」が正例となりそれ以外の候補は負例となる。ここで(1)における「友達」と(2)における「私」に着目すると、これらの候補は負例のなかでも別の文脈では正例となりうる候補である。その意味でこれらの負例は「よい」負例であり、**ニアミス候補**であるといえる。ニアミス候補は他の負例にくらべて「よい」負例であるが、述語項構造解析を2値分類タスクとして定式化するには単なる負例として扱われる。本稿ではニアミス候補を活用するため、ランキング学習の枠組みを利用した述語項構造解析器を構築する。

アノテーション時の視線を利用した研究として、述語項関係アノテーションにおけるアノテーション作業員間の不一致の検出に視線情報が寄与する [13] という報告はあるものの、これまで述語項構造解析に視線情報を利用した研究は報告されていない。本稿ではアノテーション作業員が頻繁に見る候補が「よい」負例であると仮定し、アノテーション作業員の視線データをもとに項候補を順位づけ、ランキング SVM [14] によってモデルを訓練する。

本稿の構成は以下のとおりである。2章で関連研究を外観し、3章ではタスク設定を定義し視線情報を利用した述語項構造解析手法を提案する。4章では評価実験の詳細と結果を記したうえで考察を行い、5章で結論をまとめる。

2. 関連研究

述語項構造解析の基本的な素性は、英語での研究において提案された [7]。主語の省略など日本語において特徴的な現象に対応する必要があるため、英語において有用な素性だけでは不十分であり、ゼロ照応解析などの研究を通して様々な素性が提案されてきた [8], [15], [16]。

過去の日本語述語項構造解析の手法は、大きく分けて次の3つの類型 (i) 格ごとに個別のモデルを構築し、述語ごと・格ごとに項を推定する手法 [5], [17], [18] (ii) 1つの述語に対して必須格を同時に推定する手法 [16], [19], [20] (iii) 各文のすべての述語に対して必須格を同時に推定する手法 [6], [21] に分類できる。本研究においては、省略の生じやすいガ格に焦点をあてることとしたため、本稿での提案手法は (i) に属している。

3. 視線情報の述語項構造解析への導入

3.1 タスク設定

本研究における課題は、与えられた文章の中の指定された述語に対するガ格を同定することである。日本語においては他の格とくらべてガ格が省略されやすいため、特にガ格に焦点をあてることにした。ここでの述語には事態性名詞 [8] も含まれる。また、対象の述語とガ格が同一文内に現れる課題を**文内ガ格課題**、そうでない課題を**文間ガ格課題**として区別して扱う。

この課題を解く解析器は、入力文章において対象の述語とその述語に先行する項候補が与えられたとき、正しいガ格の項を出力する。本研究の目的は、視線情報が述語項構造解析モデルの訓練に有用であることを示す点にある。

3.2 停留の検出

記録された注視点列から停留を検出する際には Dispersion-Threshold Identification アルゴリズム [22] を利用する。文章中の単語に対する停留は、その単語に対する認知過程と関連があると考えられている [23]。アルゴリズムの概要を以下に示す。ここでは注視点の集合を指して停留候補と呼び、停留候補のうち一定の条件を満たすものを停留として検出している。

- (1) 次の条件が満たされなくなるまで、注視点を停留候補に加え続ける。
 - その停留候補に含まれるすべての注視点が、停留候補の重心座標から距離閾値 D 以内に含まれる。
 - 停留候補を構成する注視点には計測状態がエラーのものを含まない。これらの条件が満たされなくなった直前までの停留候補を、1つの停留候補として固定する。
- (2) 注視点列の残りに対して (1) を繰り返し、次の停留候補を固定していく。
- (3) 各停留候補について、その持続時間が時間閾値 T 以上のものを停留として認定する。

本稿においては、光田ら [13] にならって距離閾値 D を 16 ピクセル、時間閾値 T を 100 ミリ秒とした。さらに停留の座標を停留に含まれる注視点の重心座標によって定義し、停留と項候補を対応づけた。これにより、アノテーション作業員が各候補をどの程度見ていたか、停留回数と停留の持続時間によって評価することが可能となる。

3.3 素性設計

素性設計に関しては、過去の研究 [24] で利用された基本的な素性を利用した。利用した素性の一覧を表 1 に示す。多くの素性は BCCWJ に付与された情報を利用している。距離素性 $d \in [0, 1]$ は、述語と項候補の間にある単語の数を

表 1 述語項構造解析に利用した素性

カテゴリ	ID	素性ラベル	概要
述語	1	語彙素	述語の語彙素
	2	語種	和語, 漢語, 外来語, あるいはそれらの複合
	3	品詞	述語の品詞
	4	活用形	述語の活用形
	5	活用型	述語の活用型
	6	原文文字列	述語の原文文字列
項	7	語彙素	項候補の語彙素
	8	語種	和語, 漢語, 外来語, あるいはそれらの複合
	9	品詞	項候補の品詞
	10	原文文字列	項候補の原文文字列
	11	後続格助詞	項候補に後続する格助詞
	12	意味カテゴリ	項候補の意味カテゴリ
述語と項	13	語彙素ペア	1 と 7 のペア
	14	距離	述語と項候補の距離
	15	文内/文間	述語と項候補が同一の文に出現しているか否か
	16	文内/文間と後続格助詞ペア	11 と 15 のペア
	17	意味カテゴリペア	述語の意味カテゴリと項候補の意味カテゴリのペア
	18	係り受け	直接係り受け, 間接係り受け, 係り受けなし のいずれか

正規化することによって生成し、意味カテゴリ素性については、日本語のシソーラスである『日本語語彙大系』[25]を利用した。文内/文間素性については、述語と項候補が同一文内にあるかで判定し、文境界はBCCWJ-DepParaPASに従った。

3.4 項候補の順位づけ

各アノテーション作業員について、各アノテーション課題が1つのランキングを構成するように停留をもとに項候補を順位づける。以下にその手順を示す。

- (1) 正解となる項をランキングの最上位に据える。
- (2) 正解を除いた候補の中で、最も停留回数の多かった候補を2位とする。
- (3) 停留回数が0であった候補をすべて同順の3位とする。
- (4) いずれにも該当しなかった候補はランキングから除外する。

以上の手順により生成した項候補のランキングを、ランキングSVM [14]の訓練データとして利用し、モデルのパラメータ推定を行う。テスト時には各項候補について項らしさを推定し、最も推定値が大きい候補を出力する。この手順において重要なのは、視線データはモデルのパラメータ推定にのみ利用され、テスト時には視線データが不要であるということである。これにより、視線データの存在しない新規の文章に対しても解析を行うことが可能となる。

4. 評価実験

4.1 データ

評価実験のための視線データとして、光田ら [26] によって収集されたものを利用する。光田らは日本語を対象とした述語項関係アノテーションを行う作業員の振る舞いを記録・収集する実験を行った。この実験におけるアノテーション課題は、アノテーションの対象となる述語が1つだけ指定された文章が提示され、マークされた項候補の中から対象述語のガ格を選択するものである。アノテーション作業員は対象述語のガ格を同定し、マウスを用いてクリックすることが求められる。図1はアノテーション画面のスクリーンショットであり、対象となる述語が青でハイライトされ、項候補は灰色でマークされている。

光田らの実験で使われた文章は、現代日本語書き言葉均衡コーパス (Balanced Corpus of Contemporary Written Japanese; BCCWJ) [11] から採取されている。BCCWJは様々な分野からサンプリングされた約17万の文書から構成される約1億語のサイズを持つコーパスであり、各文書には書誌情報、形態論情報がアノテーションされている。BCCWJは単語の分割に関して、長単位・短単位という2つの基準を設けている。大雑把に言えば、複合語を1つの語として扱うのが長単位であり、複合語を構成する語に分割するのが短単位である。光田らの実験においては、長単位を基準として項候補を認定している。

BCCWJは3つのサブコーパス (出版サブコーパス, 図

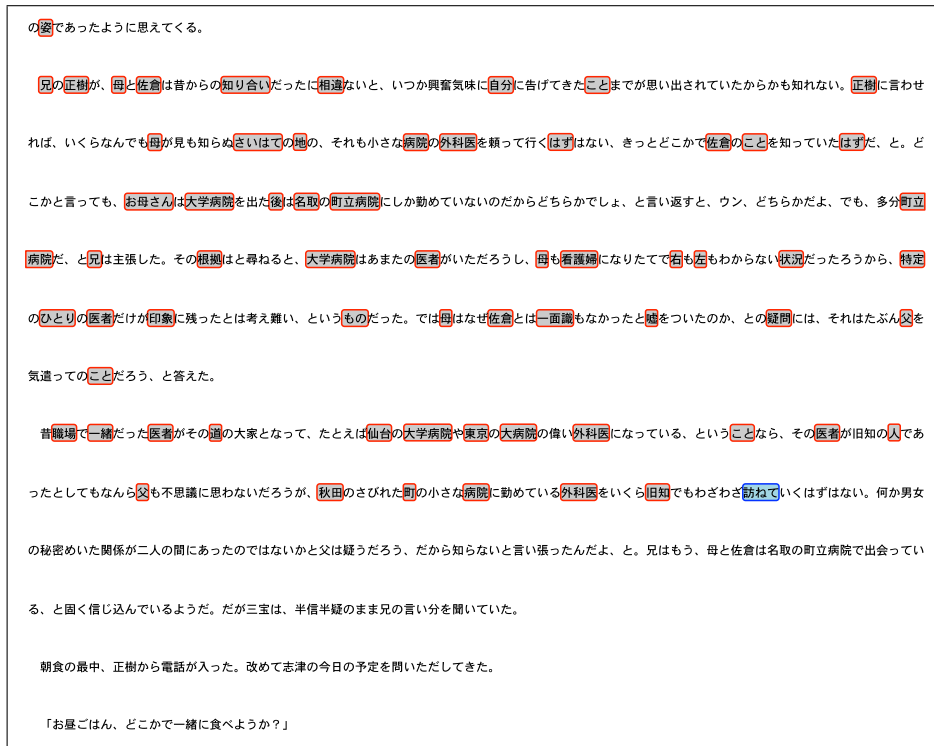


図 1 アノテーション画面のスクリーンショット

書館サブコーパス、特定目的サブコーパス) から構成されており、出版サブコーパスは書籍、新聞、雑誌の3つのレジスタから構成されている。

BCCWJの中には、機械処理されたデータに対して人手による修正を行ったサブセットがあり、コアデータと呼ばれている。コアデータに該当する文書は1980文書あり、語数にして全体の約1%に相当している。このコアデータについては係り受け構造、並列構造、照応関係、述語項構造がアノテーションされ、BCCWJ-DepParaPAS [4], [9], [10], [11], [12]として公開されている。BCCWJ-DepParaPASの述語項構造アノテーションは、NAISTテキストコーパスのガイドラインを継承し、必須格の項が短単位レベルでアノテーションされている。

光田らの実験で利用された文章はBCCWJコアデータのPBレジスタからサンプリングされたものであるため、人手で修正された形態論情報を得ることができる。またBCCWJ-DepParaPASと組み合わせることで、人手による述語項構造アノテーションの結果を利用することができる。アノテーション課題として利用された221課題のうち、光田らとBCCWJ-DepParaPASのアノテーション基準の相違により37課題を破棄することとした。残る184課題のうち、107課題が文内ガ格、77課題が文間ガ格の課題となっている。各課題における項候補の数は、正解を含めて12から109の範囲に収まり、その中央値は60であった。

光田らによる実験において、各課題は20人のアノテーション作業員によって独立にアノテーションされた。ア

ノテーションの最中、作業員の視線は視線計測装置 Tobii T60 によって 60 Hz の頻度で計測された。注視点の計測は、以前にくらべると手軽かつ正確になったとはいえ、周囲の照明やアノテーション作業員のメガネ、また作業員の目の特性など様々な要因によりエラーが生じることがある。エラーの影響を軽減するため、視線計測状況が良好な作業員に限定して視線データを利用することにした。計測状況を推し量る際には、計測エラーの多寡だけでなく停留の数もその指標となる。計測エラーが多い場合には、検出される停留の数が少なくなると考えられるためである。本稿では、セッション単位でエラー率と停留の数を計上し、それらをもとに視線を利用する作業員を限定した。ここでセッションとは1人の作業員による1課題についての計測開始から終了までの時間区間を指し、エラー率とは、各セッションにおいて60 Hzで記録された注視点の数を分母、注視点計測に失敗した回数を分子として計算されたものを指している。視線を利用する作業員は、セッション単位のエラー率が最大でも0.30を超えず、かつセッション単位の停留数が最低でも6以上という経験的な条件により限定した。以上の条件により、各課題について7人の視線データを利用することとした。

解析器の構築

今回、評価のために3種類の学習ベースのモデルとルールベースの解析器を用意した。以下では各モデルとルールベース解析器について、その詳細を記す。学習ベースモデルのための訓練データ作成には、BCCWJ-DepParaPASを

利用した。

● **2 値回帰 (BiReg) :**

2 値回帰モデルでは回帰の値が最も大きい候補をガ格の出力として扱う。各訓練事例は人手でアノテーションされたガ格のみを正例とし、残りはすべて負例とする。

● **距離順 (DRank):**

距離順モデルでは各項候補が項らしさを持つと仮定して、各項候補を順位づけする。本モデルの項らしさは、正解の項を除いて距離素性にもとづいて推定される。訓練事例は各アノテーション課題に対して用意し、ランキングは以下の手順で作成する。正解の項候補をランキングの最上位に据える。その他の項候補は距離素性の昇順に並べ、最も遠い候補を最下位とする。本モデルは視線を用いない順位づけ手法を用いることでランキングモデルの有効性を確認するために用意した。

● **停留順 (FixRank):**

停留順モデルは距離順モデルと同じく、各項候補が項らしさを持つと仮定して各項候補を順位づけする。訓練事例は各アノテーション課題につきアノテーション作業者別に作成されるため、1 課題について7つのランキングが生成される。順位づけに際しては、各項候補に対する停留の回数にもとづいて行う。その手順は3.4に記したとおりである。

● **ルールベース解析器 (Rule-base):**

ルールベース解析器では、格助詞ガによる格表示があるものを最優先し、また直接係り受けのあるものを優先したうえでガ格を予測している。具体的には以下の規則をこの順に適用し、適用できた規則にしたがって出力を決定する。

- (1) 直接の係り受け関係があり、格助詞ガにより格表示された候補を出力する
- (2) 述語が名詞であり、直接係り受けのある項候補の中に、格助詞ノにより格表示されたものがあればそれを出力する
- (3) 間接の係り受け関係があり、格助詞ガにより格表示された候補を出力する
- (4) 直接の係り受け関係がある候補のうち最も述語に近い候補を出力する
- (5) 間接の係り受け関係がある候補のうち最も述語に近い候補を出力する
- (6) 最も述語に近い候補を出力する

なお、1つの規則に対して複数の候補が該当した場合には、述語に近い候補を出力する。

テストデータ

評価のためのデータセットとして、BCCWJ コアデータの PB レジスタから作成した 29,519 課題を用意した。このうち 21,816 課題が文内ガ格の課題であり、7,703 課題が

文間ガ格の課題である。光田らによる実験では視線の分析を容易にするためスクロールが生じないよう文章量をコントロールする必要があったが、評価用の事例についてはその必要がないため、項候補数の範囲のみ合わせることにした。正解を含めた項候補数が、訓練データと同じく 12 以上 109 以下になるよう 29,519 課題を用意したところ、項候補数の中央値は 48 となった。

4.2 結果と考察

BiReg モデルの学習には Classias [27] を利用し、L2 正則化ロジスティック回帰を適用した。DRank モデルと FixRank モデルには SVM^{rank} [28] を利用した。これら2つのモデルに対しては L1 正則化項を加えたうえで学習を行った。

今回、モデルの訓練に際して、4種類の素性セットを用意した。用意した素性セットは、ベースとなる F_{base} , F_{base} に意味カテゴリ素性を追加した F_{sem} , F_{base} に係り受けの素性を追加した F_{syn} , F_{base} に意味カテゴリ素性と係り受けの素性を追加した F_{synsem} であり、それぞれに該当する素性 ID を表 2 に示す。

表 2 素性セットのバリエーション

素性セット	表 1 の素性 ID
F_{base}	1-11, 13-16
F_{sem}	1-17
F_{syn}	1-11, 13-16, 18
F_{synsem}	1-18

いずれのモデルも各項候補のガ格らしさを推定しているため、推定値が最も大きい候補をガ格として出力する。その出力がコーパスのアノテーションと一致した場合のみ「正解」と判定し、正解の項と共参照の関係にある候補であってもコーパスのアノテーションと異なっていれば不正解として扱うこととした。

表 3 は各モデルと素性セットについての評価結果を表したものである。精度の計算に際しては、文内ガ格と文間ガ格を分けて算出した。表 3 の結果から、文内・全体では FixRank+ F_{syn} モデルが最も高い精度を示しており、文間では BiReg モデルが全体的により精度を示していることがわかる。

この表から判断する限り、FixRank+ F_{syn} モデルが文内ガ格を最もよく予測している。この結果は、アノテーション作業者の各項候補に対する停留は、その候補の項らしさを反映しているという仮説を支持している。また、FixRank+ F_{syn} モデルは同一素性セットである BiReg+ F_{syn} モデルとくらべて精度が 0.05 ポイント向上しており、視線情報が述語項構造解析モデルのパラメータ推定に視線情報が寄与することもわかった。

表 3 評価結果 (精度)

モデル	素性セット	文内ガ格	文間ガ格	全体
BiReg	F_{base}	0.56	0.04	0.42
	F_{sem}	0.48	0.06	0.37
	F_{syn}	0.58	0.03	0.44
	F_{synsem}	0.52	0.05	0.40
DRank	F_{base}	0.47	0.01	0.35
	F_{sem}	0.47	0.01	0.35
	F_{syn}	0.50	0.01	0.37
	F_{synsem}	0.51	0.01	0.38
FixRank	F_{base}	0.55	0.02	0.41
	F_{sem}	0.49	0.02	0.37
	F_{syn}	0.63	0.02	0.47
	F_{synsem}	0.58	0.02	0.43
Rule-base		0.56	0.01	0.42

素性セットとして F_{sem} を利用した欄を見ると、2 値回帰モデルの文間ガ格を除いて素性が有効に機能していないことがわかる。今回、意味カテゴリを付与するにあたっては語義曖昧性解消なしに複数レベルのカテゴリを付与している。Taira ら [5] のように意味カテゴリを人手で付与することで F_{sem} が有効に働く可能性がある。

ここで DRank モデルが全モデルの中で最も悪い結果を示していることに着目する。DRank モデル用の訓練事例は、距離素性にもとづいた同順位を許さないランキングであるため、項予測の妨げとなるような過剰な制約を多分に含んでいる可能性がある。これに対して FixRank モデルは、3 段階のランキングであり同順位の候補を増やすことで学習時の制約が少なくなるようになっている。これらのモデルの特性が精度の良し悪しに影響したのだと思われる。

さらなる分析のため、(11) 後続格助詞、(15) 文内/文間、(18) 係り受けの 3 素性に焦点をあて、これらの素性にもとづき適合率、再現率、F 値を計算した。表 4 は BiReg+ F_{syn} と FixRank+ F_{syn} の両モデルの比較である。

両者のモデルを比べてみると、直接係り受けのあるガ格については FixRank モデルの方がより高い F 値を示しており、特に格助詞がで格表示されたガ格についてはその傾向が顕著である。直接係り受けがあるガ格に対する FixRank モデルによる項予測は、BiReg モデルより高い再現率を示している一方で、格助詞がで格表示されていないガ格に対しては低い適合率を示している。両者のモデルについて適合率を計算する際の事例数、すなわち直接係り受けのある候補をガ格として予測した事例数とその候補が実際にガ格であった事例数を見てみたところ、FixRank モデルでは予測した 19,612 事例中 11,095 事例が、BiReg モデルでは予測した 12,520 事例中 8,738 事例が、実際に正解のガ格であった。このことは FixRank モデルが、直接係り

受けのある候補をガ格として予測する傾向にあることを示している。

このことに関連して、FixRank モデルの訓練データにおいて第 2 位の項候補が直接係り受けの関係にある事例数を調べてみたところ、1,288 事例 (= 184 課題 × 7 人) のうち 428 事例が該当し、全訓練事例の 33 % に相当した。一方で正解の候補については 184 課題のうち 35 課題が述語と直接係り受けの関係にあり、全訓練事例の 19 % であった。これらの結果は、アノテーション作業者が直接係り受けのある候補をアノテーション時によく見ており、FixRank モデルがそれらの「よい」負例をガ格として予測する傾向にあることを示している。FixRank モデルが文間ガ格を予測できていないのはこの傾向のためだと考えられる。

5. 結論

本稿では日本語述語項構造解析において視線情報を取り入れる手法を提案した。対象の述語に対してガ格を予測する評価を通じて、視線情報が述語項構造解析に有効であることを確認した。特に、視線情報をパラメータ推定に導入することで精度を 0.05 ポイント向上させるという結果を得た。しかしながら、文間ガ格に対する精度は依然としてとても低い結果となっている。この点を考えると、さらなる改善のためには視線情報の使い方を見直す必要があると考えられる。現在のところ、視線情報として利用しているのは停留の回数のみであり、停留の持続時間や時系列の情報は利用していない。これらの情報を用いることで新たな方向性が見えてくる可能性がある。

今回の評価において DRank モデルをランキングモデルのベースラインとしたが、その性能は低く、適切なベースラインとして認めがたい結果であった。さらに適切な評価を行うためにはパラメータ推定に際してランキング学習を行っている萩行ら [29] の手法を取り入れる必要があると考えている。

視線情報は、近年さまざまな自然言語処理分野で注目されており、対話システム [30], [31], 照応解析 [32], 統語解析 [33], 共参照解析 [34], 固有表現抽出 [35] での取り組みが報告されている。他の自然言語処理タスクにおいても、アノテーション時の視線情報の有効性を確かめることは研究の方向づけに新たな視点をもたらす可能性がある。

参考文献

- [1] Pustejovsky, J. and Stubbs, A.: *Natural Language Annotation for Machine Learning*, O'Reilly (2012).
- [2] Tokunaga, T., Iida, R. and Mitsuda, K.: Annotation for annotation – Toward eliciting implicit linguistic knowledge through annotation –, *Proceedings of the 9th Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation (ISA-9)*, pp. 79–83 (2013).
- [3] 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

表 4 BiReg+ F_{syn} モデルと FixRank+ F_{syn} モデルの比較

(15) 文内/文間	(18) 係り受け	(11) 後続格助詞	BiReg+ F_{syn}			FixRank+ F_{syn}		
			P	R	F	P	R	F
文内	直接	ガ	0.949	0.884	0.915	0.948	0.975	0.962
		その他	0.513	0.463	0.487	0.403	0.693	0.509
	間接	ガ	0.227	0.602	0.330	0.242	0.561	0.338
		その他	0.184	0.166	0.175	0.156	0.103	0.124
なし	ガ	0.294	0.892	0.442	0.352	0.741	0.477	
	その他	0.239	0.469	0.316	0.306	0.290	0.297	
文間	なし	ガ	0.232	0.050	0.082	0.284	0.025	0.046
		その他	0.247	0.031	0.055	0.172	0.014	0.025

「その他」には格助詞が付随しない場合も含まれる。

P, R, F はそれぞれ適合率, 再現率, F 値を表す。

- 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).
- [4] 植田禎子, 飯田 龍, 浅原正幸, 松本裕治, 徳永健伸: 『現代日本語書き言葉均衡コーパス』に対する述語項構造・アノテーション, 第 8 回コーパス日本語学ワークショップ予稿集, 国立国語研究所, pp. 205–214 (2015).
- [5] Taira, H., Fujita, S. and Nagata, M.: A Japanese Predicate Argument Structure Analysis Using Decision Lists, *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2008)*, pp. 523–532 (2008).
- [6] Shibata, T., Kawahara, D. and Kurohashi, S.: Neural Network-Based Model for Japanese Predicate Argument Structure Analysis, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 1235–1244 (2016).
- [7] Gildea, D. and Jurafsky, D.: Automatic Labeling of Semantic Roles, *Computational Linguistics*, Vol. 28, No. 3, pp. 245–288 (2002).
- [8] Komachi, M., Iida, R., Inui, K. and Matsumoto, Y.: Learning-Based Argument Structure Analysis of Event-Nouns in Japanese, *Proceedings of the Conference of the Pacific Association for Computational Linguistics*, pp. 120–128 (2007).
- [9] 浅原正幸, 大村 舞: BCCWJ-DepParaPAS: 『現代日本語書き言葉均衡コーパス』係り受け・並列構造と述語項構造・共参照アノテーションの重ね合わせと可視化, 言語処理学会第 22 回年次大会発表論文集, pp. 489–492 (2016).
- [10] 小西 光, 小山田由紀, 浅原正幸, 柏野和佳子, 前川喜久雄: BCCWJ 係り受け関係アノテーション付与のための文境界再認定, 第 3 回コーパス日本語学ワークショップ予稿集, 国立国語研究所, pp. 135–142 (2013).
- [11] Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koiso, H., Yamaguchi, M., Tanaka, M. and Den, Y.: Balanced corpus of contemporary written Japanese, *Language resources and evaluation*, Vol. 48, No. 2, pp. 345–371 (2014).
- [12] 浅原正幸, 松本裕治: 『現代日本語書き言葉均衡コーパス』に対する係り受け・並列構造アノテーション, 言語処理学会第 19 回年次大会発表論文集, pp. 66–69 (2013).
- [13] Mitsuda, K., Iida, R. and Tokunaga, T.: Detecting Missing Annotation Disagreement using Eye Gaze Information, *Proceedings of the 11th Workshop on Asian Language Resources*, pp. 19–26 (2013).
- [14] Joachims, T.: Optimizing search engines using click-through data, *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD 2002)*, pp. 133–142 (2002).
- [15] Iida, R., Inui, K. and Matsumoto, Y.: Zero-anaphora Resolution by Learning Rich Syntactic Pattern Features, *ACM Transactions on Asian Language Information Processing (TALIP)*, Vol. 6, No. 4, pp. 1:1–1:22 (2007).
- [16] Sasano, R. and Kurohashi, S.: A Discriminative Approach to Japanese Zero Anaphora Resolution with Large-scale Lexicalized Case Frames, *Proceedings of 5th International Joint Conference on Natural Language Processing*, pp. 758–766 (2011).
- [17] Imamura, K., Saito, K. and Izumi, T.: Discriminative Approach to Predicate-argument Structure Analysis with Zero-anaphora Resolution, *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pp. 85–88 (2009).
- [18] Hayashibe, Y., Komachi, M. and Matsumoto, Y.: Japanese Predicate Argument Structure Analysis Exploiting Argument Position and Type, *Proceedings of 5th International Joint Conference on Natural Language Processing*, pp. 201–209 (2011).
- [19] Yoshikawa, K., Asahara, M. and Matsumoto, Y.: Jointly Extracting Japanese Predicate-Argument Relation with Markov Logic, *Proceedings of 5th International Joint Conference on Natural Language Processing*, pp. 1125–1133 (2011).
- [20] Hangyo, M., Kawahara, D. and Kurohashi, S.: Japanese Zero Reference Resolution Considering Exophora and Author/Reader Mentions, *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 924–934 (2013).
- [21] Ouchi, H., Shindo, H., Duh, K. and Matsumoto, Y.: Joint case argument identification for Japanese predicate argument structure analysis, *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 961–970 (2015).
- [22] Salvucci, D. D. and Goldberg, J. H.: Identifying Fixations and Saccades in Eye-tracking Protocols, *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA 2000)*, pp. 71–78 (2000).
- [23] Just, M. A. and Carpenter, P. A.: A Theory of Read-

- ing: From Eye Fixations to Comprehension, *Psychological Review*, Vol. 87, No. 4, pp. 329–354 (1980).
- [24] 松林優一郎, 乾健太郎: 統計的日本語述語項構造解析のための素性設計再考, 言語処理学会第20回年次大会発表論文集, pp. 360–363 (2014).
- [25] 池原 悟, 宮崎正弘, 白井 諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林 良彦: 日本語語彙大系 CD-ROM 版, 岩波書店 (1999).
- [26] 光田 航, 飯田 龍, 徳永健伸: 単一述語項関係アノテーション課題における視線情報の収集と分析, 情報処理学会第217回自然言語処理研究会, pp. 1–8 (2014).
- [27] Okazaki, N.: Classias: a collection of machine-learning algorithms for classification (2009).
<http://www.chokkan.org/software/classias/>.
- [28] Joachims, T.: Training Linear SVMs in Linear Time, *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)* (2006).
- [29] 萩行正嗣, 河原大輔, 黒橋禎夫: 外界照応および著者・読者表現を考慮した日本語ゼロ照応解析, 自然言語処理, Vol. 21, No. 3, pp. 563–600 (2014).
- [30] Prasov, Z., Chai, J. Y. and Jeong, H.: Eye Gaze for Attention Prediction in Multimodal Human-Machine Conversation, *Proceedings of the AAAI Spring Symposium on Interaction Challenges for Artificial Assistants*, pp. 102–110 (2007).
- [31] Qu, S. and Chai, J. Y.: An exploration of eye gaze in spoken language processing for multimodal conversational interfaces, *Proceedings of Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 284–291 (2007).
- [32] Prasov, Z. and Chai, J. Y.: What’s in a gaze?: The role of eye-gaze in reference resolution in multimodal conversational interfaces, *Proceedings of the 13th international conference on Intelligent user interfaces*, pp. 20–29 (2008).
- [33] Barrett, M. and Søgaard, A.: Using reading behavior to predict grammatical functions, *Proceedings of the Sixth Workshop on Cognitive Aspects of Computational Language Learning*, pp. 1–5 (2015).
- [34] Ross, J. C., Mishra, A. and Bhattacharyya, P.: Leveraging annotators’ gaze behaviour for coreference resolution, *Proceedings of the 7th Workshop on Cognitive Aspects of Computational Language Learning*, pp. 22–26 (2016).
- [35] Tomanek, K., Hahn, U., Lohmann, S. and Ziegler, J.: A Cognitive cost model of annotations based on eye-tracking data, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*, pp. 1158–1167 (2010).