修士論文

BERTに基づく統合的日本語結束性解析

指導教員　　黒橋 禎夫 教授

京都大学大学院情報学研究科
修士課程知能情報学専攻

植田 暢大

令和3年2月1日

BERT に基づく統合的日本語結束性解析

植田 暢大

内容梗概

自然言語で書かれたテキストの背後には種々のエンティティ間の結束性が存在し，これによってテキストの意味が支えられている．このような結束性には，述語項構造，橋渡し照応関係，共参照関係などが含まれる．述語項構造は,「誰が何を何にどうした」というように，中心となるある述語に加えてその述語の主語や目的語に相当する項から構成される構造である．通常，述語項構造は動詞や形容詞などに対し議論されるが，本研究では一部の名詞に対しても項構造を考え，同様に扱う．橋渡し照応関係は，典型的には「A の B」という形で結ばれるような名詞 A, B の間に成り立つ関係で，特に A が B にとって必須的な情報を補っている場合を指す．共参照関係は，実世界において同一であるエンティティを参照する名詞間の関係である．これらの関係を認識することは，コンピュータが自然言語を理解するための基盤となる．

一般に，テキスト中ではこれらの関係が互いに強く結びついており，その間には相互作用が存在する．その相互作用を捉えるには，これら関係を統合的に解析することが重要である．しかし，名詞に対する述語項構造や橋渡し照応関係の解析は非常に困難であり，これまで統合的な解析に取り組んだ研究は存在しない．しかし，近年のニューラルネットワークの進歩，特に BERT に代表される自己教師あり学習を利用した事前学習モデルは，多くの自然言語処理タスクを大幅に改善した．これらの技術を用いることで，名詞に対する述語項構造や橋渡し照応関係を含む文章全体の結束性の研究も視野に入れることが可能になってきた．そこで，本研究では BERT に基づき，日本語のテキストにおける結束性の統合的な解析に取り組む.

実験の結果，本研究による手法はいずれのタスクにおいても既存手法を大幅に上回った．また，様々な組み合わせで同時解析を行うことで，共参照関係がその他の関係と本質的に異なることを確認し，解析においては特別に扱うべきであることを示した.

BERT-based Cohesion Analysis of Japanese Texts

Nobuhiro UEDA

Abstract

The meaning of natural language text is supported by cohesion among various kinds of entities, including coreference relations, predicate-argument structures, and bridging anaphora relations. A predicate-argument structure (PAS) consists of a predicate and its arguments that fill each case of the predicate, such as *who* does/did *what* to *whom*. Although PAS is usually considered for verbal predicates, such as verbs and adjectives, this study also considers PAS analysis for nominal predicates, which is syntactically noun but takes arguments. A bridging anaphora relation is an anaphoric relation between an anaphor and its antecedent, where the antecedent makes up a kind of semantic insufficiency of the anaphor. A coreference relation is a relation between nouns that refer to the same real-world entity. Recognizing these relations is the basis for computers to understand natural language texts.

In general, these relations are strongly connected to each other, and there are interactions among them. In order to capture the interactions, it is important to analyze these relations in an integrated manner. However, it is very difficult to analyze nominal predicate-argument structures and bridging anaphora relations, and there are no studies on the integrated analysis of the relations. Recent advances in neural networks, in particular self training-based language models including BERT, have significantly improved many natural language processing (NLP) tasks. Using these techniques, it is now possible to dive into the study on analysis of cohesion in the whole text, including nominal predicate-argument structures and bridging anaphora relations. In this study, we tackle an integrated analysis of cohesion in Japanese texts based on BERT.

Our results significantly outperformed existing studies in each task. Furthermore, we also showed that coreference relations are different in nature from the other relations and should be treated specially.

BERT に基づく統合的日本語結束性解析

目次

第 1 章　はじめに　1

第 2 章　関連研究　7

2.1　日本語結束性解析 7

2.2　英語結束性解析 9

2.3　BERT 10

2.3.1　Transformer 10

2.3.2　入力表現 11

2.3.3　Pre-training 12

2.3.4　Fine-tuning 13

第 3 章　日本語結束性タグ付きコーパス　14

3.1　各コーパスの特徴 14

3.2　作業者間一致度 15

第 4 章　BERT に基づく結束性の統合的解析手法　17

4.1　手法の概要 17

4.2　解析対象基本句の選択 17

4.3　同時解析モデル 20

4.3.1　入力系列 21

4.3.2　出力層 22

4.4　共参照解析を特別に扱ったモデル 23

4.5　係り受けを考慮したモデル 25

第 5 章　実験　27

5.1　実験設定 27

5.2　評価手法 30

5.3　CAModel 31

5.3.1　概要 31

5.3.2　結果 31

5.4　CorefCAModel 35

5.4.1 概要 35
5.4.2 結果 35
5.5 DepCAModel 36
5.5.1 概要 36
5.5.2 結果 38
5.6 議論 38
第 6 章 おわりに **43**
謝辞 **44**
参考文献 **45**

第1章　はじめに

自然言語で書かれたテキストの背後には種々のエンティティ間の結束性 (cohesion) が存在し，これによってテキストの意味が支えられている．このような結束性には，述語項構造，橋渡し照応関係，共参照関係などが含まれる．例として，これらの関係を図 1 に示す．これらの関係を認識することは，自然言語テキストの意味的構造を捉えることに他ならず，コンピュータが自然言語を理解するための基盤となる．

これらの関係は図に示すように複数の単語を共有し，互いに強く結びついている．したがって，これら関係には相互作用があると考えられ，その相互作用を捉えるためこれら関係は統合的に解析することが重要である．しかし，このような統合的なテキストの結束性の分析に焦点を当てた研究はこれまで存在しない．その一因として，タスク単独の難しさが挙げられる．これまで共参照関係や動詞に対する述語項構造に関する研究は盛んに行われてきた一方，名詞に対する述語項構造や橋渡し照応関係の解析は非常に困難であり，あまり研究されてこなかった．しかし，近年のニューラルネットワークの進歩，特に BERT [1] に代表される自己教師あり学習を利用した事前学習モデルは，多くの自然言語処理タスクを大幅に改善した．これらの技術を用いることで，名詞に対する述語項構造や橋渡し照応関係を含む文章全体の結束性の研究も視野に入れることが可能になってきた．そこで本研究では，日本語のテキストにおける結束性の統合的な解析に取り組む．

統合的な学習では関連のあるタスク同士に利得があることが知られている [2, 3, 4]．そこで，本研究が対象とする結束性を構成する関係についてその特徴を整理する．

述語項構造は，述語とその述語の格を満たす項から構成される．項とは一般に対象の述語にとって必須的な情報を補う役割を持つ単語であり，例えば「誰が」や「何を」に相当する単語が主要な項として挙げられる．また，格とは述語に対する項の役割を指し,「誰が」や「何を」に相当する役割はそれぞれガ格, ヲ格などとよばれる．述語に対してこれら項を特定するタスクを述語項構造解析とよぶ．

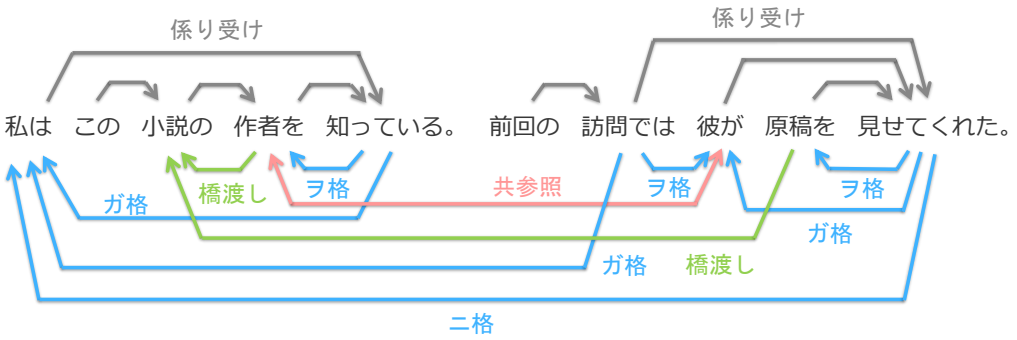

図 1: 結束性を構成する意味的関係の例.

表 1: 述語項構造解析の分類.

述語項構造解析	*overt* についての解析	
	case についての解析（格解析）	
	zero についての解析（ゼロ照応解析）	文脈照応解析
		外界照応解析

述語項構造解析の分類を表 1 に示す．述語項構造における述語と項はそれらの関係によって *overt*, *case*, *zero* の 3 種類に分類され，それぞれ解析タスクの難易度が異なる．*overt* とは，図 1 の「作者を」と「知っている」のように，項が「が」,「を」,「に」などのケースマーカーを伴い，かつ述語と係り受けの関係を持っている場合を指す．*overt* の場合，係り受け関係さえ正しく解析できれば述語項構造を決定することができる．近年，構文解析は 92%以上の精度が達成されており [5]，*overt* は高精度で解析が可能である．一方，項と述語が係り受け関係を持っていても，項に「は」や「も」などの主題を表す助詞が付随する場合，格は表層から決定できない．例えば，図 1 の 「私は」と「知っている」はガ格の関係にあるが，それは文中に明示されていない．このような場合を *case* とよび，*case* の関係にある述語と項の解析を格解析とよぶ．*overt* あるいは *case* は述語と項が係り受けの関係にある場合の分類だが，述語に対して項が省略されている場合，述語と項は係り受けの関係を持たない．このとき，述語と項の関係を特に *zero* とよび，その解析をゼロ照応解析とよぶ．例えば，図 1 の 2 文目では,「見せてくれた」のニ格の項が省略されており，1 文目の「私」を参照

する必要がある．

表 1 に示すように，ゼロ照応解析はさらに文脈照応解析と外界照応解析に分類される．文脈照応解析は,「見せてくれた」に対する「私」のように解析対象がテキスト中に出現している場合を指す．一方，以下の例では述語「休憩する」に対するガ格の項がテキスト中に出現しない．

(1)　疲れたので少し（ϕ ガ）休憩する。[1)]

このような場合を，外界照応解析とよぶ．一般に外界照応解析では項を一意に定めることができない．上記の例では例えば「私」が項として考えられるが，その他にも「僕」や「自分」など無数の可能性が存在する．したがって，外界照応解析では「私」などの具体的な語ではなく［著者］などのように汎化した表現を定め，これを解析の対象とするのが一般的である．

述語項構造解析は通常，動詞や形容詞などの用言に対して行われる．この場合を用言述語項構造解析とよぶ．しかし，名詞にもサ変名詞をはじめとして用言のように項を取るものが多く存在する．このような名詞は事態性名詞とよばれる．本研究では，事象性名詞に対する述語項構造解析も考慮し，これを体言述語項構造解析とよぶ．体言述語項構造解析において *overt* の関係はほとんど存在しない[2)].「原稿執筆」や「原稿の執筆」における「原稿」と「執筆」は *case* の関係だが，体言述語項構造解析ではほとんどの場合，図 1 の「訪問」のように述語と項との間に係り受け関係がないため，ゼロ照応解析を行う必要がある．

橋渡し照応とは間接照応ともよばれ，照応現象の一つである．そこでまず照応について説明する．照応とはテキスト中である表現が他の表現を参照するという現象である．このとき，他を参照する表現を照応詞，照応詞に参照される表現を先行詞とよぶ.

表 2 に示すように，照応は大きく，直接照応と間接照応，さらにそれぞれゼロ照応かそうでないかに分類される．なお，ここではゼロ照応でない照応を非ゼロ照応とよぶ．文脈照応か外界照応かという分類も存在するが，それについては先に説明した通りである．照応現象の中でも最も典型的なものが直接非ゼロ照応である．直接非ゼロ照応はテキスト中の代名詞などのある表現が同一の

[1)] ゼロ照応解析において省略された項はゼロ代名詞とよばれ，ϕ で表される．

[2)] 例えば,「再配送はセンターに連絡後、一時間ほどで来てくださいました。」における名詞「連絡」に対する項「センター」のように，ごくわずかだが例外が存在する．

表 2: 照応現象の分類．照応詞を太字で，先行詞を下線で示している．

	非ゼロ照応	ゼロ照応
直接照応	太郎は本を取り出して、 **それ**を読んだ。	太郎はパンを買った。 （ϕ **ガ**）（ϕ **ヲ**）家で食べた。
間接照応	遠くに家が見えた。 **屋根**に雪が積もっていた。	コンビニに行って、 私は（ϕ **ニ**）道を尋ねた。

内容の他の表現を参照する現象である．表では，照応詞「それ」が同じ内容である「本」を参照している．本研究ではこれを後述する共参照解析として扱う．次に，直接ゼロ照応は直接照応の中でも照応詞が省略されている場合である．直接ゼロ照応解析ではまず，照応詞が省略されていることを検出する必要があるが，その際に述語の必須格が埋められているかどうかが手がかりとして使用される．ゼロ照応解析が述語項構造解析の部分問題となっているのはこのためである．

続いて，間接照応について説明する．間接照応とは照応詞が先行詞を参照しているが，それらが同一の内容ではない場合を指す．表 2 の間接非ゼロ照応の例では,「屋根」は「家の屋根」という関係で「家」を参照しているが，これらは同一の内容ではない．本研究ではこの間接非ゼロ照応を橋渡し照応とよび，橋渡し照応解析として解析の対象とする．橋渡し照応解析は，照応詞と先行詞の関係が全体・部分や所有をはじめ多岐にわたるため，直接照応に比べて非常に難しいタスクである．間接ゼロ照応では照応詞と先行詞の内容が異なり，かつ照応詞がテキスト中に出現しない．表 2 の例において，述語「尋ねた」のニ格の項はコンビニの従業員と考えられるが，その項は省略されており，さらに同一の表現も存在しない．この解析は難しいとされる間接照応解析の中でも特に困難である．そのため本研究では，外界照応の照応先として［不特定: 人］という対象を定め,「尋ねた」のニ格を外界（［不特定: 人］）への直接照応として扱う．このとき,［不特定: 人］が「コンビニ」の従業員であるということは解析の対象としない．

一般に，照応解析は照応詞において単独では意味的に欠けている情報を，先行詞を予測することで補完するタスクと考えられる．このとき，補完される情報の量は先行詞によって様々である．本研究では，照応詞に対して先行詞が補

うことのできる情報量を先行詞の必須性と定義し，橋渡し照応解析においては必須性の高い先行詞のみを予測する．例えば，以下の例で照応詞「記録」に対する先行詞を予測することを考える．

(2)　彼の平泳ぎの世界**記録**

このとき,「彼の記録」,「平泳ぎの記録」,「世界の記録」という意味関係がいずれも成り立ち,「彼」,「平泳ぎ」,「世界」は全て「記録」の先行詞であると考えられる．しかし，本研究では「平泳ぎ」のみを先行詞として解析する．これは照応詞「記録」にとって，最も不足している情報はその「競技」であり,「平泳ぎ」の必須性が最も高いと考えられるからである．また，橋渡し照応解析は「記録」のような単独では意味的に不十分な名詞に対して行われる．このような名詞は不飽和名詞とよばれ，他には「屋根」や「価格」などが挙げられる.「屋根」は何らかの建物の屋根であり,「価格」は何らかの商品の価格である．図 1 では,「作者」の先行詞は「小説」であり,「原稿」の先行詞もまた「小説」である．一方で，上記の例の「平泳ぎ」のように単独で意味的に充足している名詞に対しては橋渡し照応解析は行われない．

共参照関係とは，実世界における同一の実体を参照する名詞同士の関係であり，この関係を明らかにするタスクは共参照解析とよばれる．図 1 において,「作者」と「彼」は，どちらも実世界では同一であるものを異なる表現で参照している．共参照解析において，実体を参照する表現をメンションとよぶ．共参照関係は先に述べた直接非ゼロ照応関係と似通ったものであるが，わずかに違いも存在する．先に述べた直接非ゼロ照応の関係にある語のペアは，多くの場合共参照関係にある．例外としては以下のような関係が挙げられる．

(3)　太郎はプリウスを買った。次郎も**その車**を買った。（[6] より）

この場合,「その車」は「プリウス」と照応関係にあるが，両者は同一の実体ではない．しかし，こういった例は稀であるため，本研究では直接非ゼロ照応解析は扱わず共参照解析のみを対象とする．

以上，本研究で扱う 4 つのタスク，用言述語項構造解析・体言述語項構造解析・橋渡し照応解析・共参照解析を整理したが，これらタスクの関係を考えると用言述語項構造解析と体言述語項構造解析はほとんど同様の性質を持ち，橋

渡し照応解析も対象が名詞ではあるがエンティティ間のある種の必須的な関係を捉えるタスクなので用言述語項構造解析，体言述語項構造解析に近い．一方で，共参照は同一のエンティティ間に成り立つ関係であり，さらにそこに必須性は関係しない．したがって，共参照解析は用言述語項構造解析，体言述語項構造解析，橋渡し照応解析とは根本的に性質が異なる．本研究では，この性質の違いを考慮しつつ，これらの文章中のエンティティ間の関係を統合的に解析するモデルを提案し，実験を通して議論を行う．

第2章　関連研究

本章ではまず日本語における結束性解析の取り組みについて説明し，続いて英語における解析について取り上げる．さらに，結束性解析を含む様々な自然言語処理タスクで最高精度を更新した BERT [1] と呼ばれる事前学習モデルについて説明する．

2.1　日本語結束性解析

本研究で扱う 4 つの意味的関係解析タスクの中でも，日本語用言述語項構造解析は比較的易しく学習データが豊富に存在することから盛んに研究されている [7, 8, 9, 10, 11, 12, 13, 14]．Ouchi ら [13] は Grid-RNN という機構を導入することで複数の述語間の関係を捉えることを試み，その有効性を示した．Matsubayashi and Inui [11] は，述語間の関係を捉える際，RNN の代わりに attention 機構を採用することで，語順や距離に依存しない述語間関係を捉えることを試み，Ouchi らに対する優位性を示した．Kurita ら [9] は，日本語の述語項構造タグ付きコーパスのサイズ不足に着目し，敵対的学習を用いて生コーパスを活用する手法を提案した．本研究では，Kurita らが使用したコーパスに加え，異なる 2 種類のコーパスを使用することでデータ不足の問題を補っている．また，本研究で使用する BERT は後述する pre-training において大規模な生コーパスを使用しており，タグ付きコーパスの学習データに含まれない低頻度語に対しても汎化が可能と考えられる．

日本語用言述語項構造解析では，かつては人手で整備した特徴量と機械学習に基づく手法が主流だったが，このように近年はニューラルネットワークを使用したモデルが多く提案され，人手で特徴量を整備する必要がなくなった．しかしながら，ニューラルネットワークを使用したモデルにおいても，基礎解析から得られた多くの素性を加えているものがほとんどである．Matsubayashi ら [12] の手法では品詞・活用形・文節内の位置・係り受けなどの様々な素性が使用されている．このような手法は前段の基礎解析器による解析誤りの影響を受けてしまうという問題がある．それに比べ，本研究の手法は述語の選択において素性が使用されるものの，その項を予測する際には一切の素性を必要としない．

日本語用言述語項構造解析においては多くの研究 [7, 10, 11, 12] が，1 文あるいは 1 文書の入力に対してその中の 1 つの述語を解析するというモデルを提案

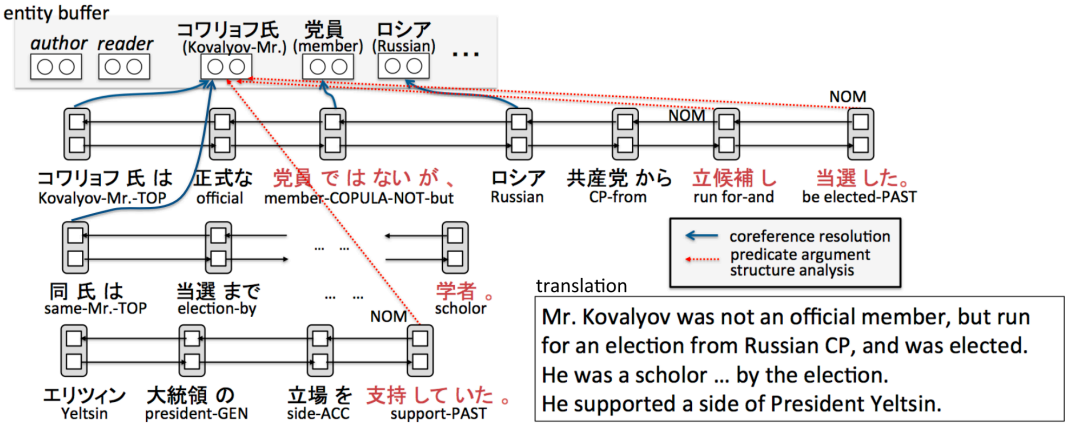

図 2: Shibata らの手法の概略（[8] より）.

している．一方，本研究では 1 つの入力文書からその文書に含まれる全ての述語についてその項を解析することが可能である．したがって，本研究は既存研究と比べ，計算効率の面でも優位性がある．

以上の研究はいずれも RNN ベースのモデルを使用しているが，本研究では self-attention 機構に基づいた BERT [1] を使用する．self-attention 機構は RNN に比べ，入力文書中の長距離の依存関係を捉えることができる．また，BERT を使用することで， pre-training を通して学習された言語の普遍的な特徴を利用することができる．

以上のように用言述語項構造解析に取り組んだ研究は多く存在するが，その他のタスクを扱った研究や複数タスクの同時解析を行った研究は多くない．Shibata ら [8] は，*entity buffer* とよばれる機構を導入し，用言述語項構造解析と共参照解析の同時学習を行った．図 2 に手法の概略を示す．図に示すように，*entity buffer* とは解析時に発見したエンティティをベクトルとして保存しておくもので，用言述語項構造解析や共参照解析を解く過程でその情報が使用され，同時に更新される．この機構により，入力文書において焦点の当てられているエンティティをモデルが捉えることが可能となる．彼らは用言述語項構造解析と共参照解析が異種のタスクであるにも関わらず，用言述語項構造解析を行うことで共参照解析の精度が向上することを示した．これは，用言述語項構造解析と共参照解析がいずれも焦点の当てられているエンティティを参照しやすいという共通の性質を持つためだと考えられる．Omori ら [7] は，用言述語項構造解析だけでなく体言述語項構造解析も同時に行い，双方に利得があることを示した．しか

し，いずれの研究も共参照解析と体言述語項構造解析の両方を行ってはいない．本研究では用言述語項構造解析，体言述語項構造解析，共参照解析に加え，橋渡し照応解析も同時に扱う．

日本語橋渡し照応解析および共参照解析についての先行研究は多くない．橋渡し照応解析について，Sasano ら [15] は大規模コーパスから「A の B」のような表現を集めて辞書を構築し，この辞書に基づく解析手法を提案した．共参照解析についても同様の手法が提案されており，Sasano ら [16] は同義表現を生コーパスから自動獲得し，その結果を解析に使用した．また，井之上ら [17] は品詞や係り受けなどの特徴量から機械学習モデルにより日本語橋渡し照応解析および直接照応解析を行った．これら先行研究はいずれもニューラルネットワークは使用しておらず，解析可能な対象は限定的であると考えられる．

2.2 英語結束性解析

日本語では格助詞が存在するため，述語の格としてはガ・ヲ・ニなどの表層格が解析の対象となることが多い．しかし，格助詞を持たない英語では表層格を定めることが難しく,「動作主」,「対象」,「道具」などの深層格が主な解析対象となる．例えば，以下の例文では述語 “broke” に対して動作主，対象，道具はそれぞれ “John”, “window”, “hammer” である．

(4) John broke the window with a hammer

この深層格を解析するタスクを意味役割付与 (Semantic Role Labeling, SRL) とよび，日本語における述語項構造解析とおおよそ対応する．SRL では BERT [1] を改良した RoBERTa [18] を使用することで F1 スコア 88.03 が達成されている [19]．また，構文解析などの類似したタスクとの同時学習が効果的であることが示されている [20].

日本語述語項構造解析と英語 SRL を比較すると，それぞれに特有の難しさが存在する．SRL では表層から格が決定できないため，英語 SRL のほうが格を決定するのは難しい．しかし，英語では一度出現した単語を再度使用する場合，代名詞で置き換えられる場合が多い一方で，日本語ではそれが省略される傾向にある．したがって，項の選択においては日本語述語項構造解析は英語 SRL より格段に難しいタスクといえる．Shibata ら [8] の研究では，格解析は F1 スコ

ア 89.5% で解けるのに対し，ゼロ照応解析は 58.1% に留まっている．

次に，英語における橋渡し照応解析・共参照解析に関する研究を取り上げる．英語においてこれら解析は，係り受けを基本とした日本語の解析とは異なり，スパンを基本として行われる [21, 22, 23]．これは，重要語が名詞句の末尾に位置するという性質を持つ日本語ではその重要語のみを予測できれば十分である一方，英語では名詞句中の重要語の位置が不定だからである．英語タスクのこの性質上，本研究と直接の比較はできないが，Hou ら [21] や Wu ら [22] は質問応答の枠組みを使用することで，橋渡し照応解析・共参照解析の精度をそれぞれ向上させた．Yu ら [23] は橋渡し照応解析における共参照解析との同時学習の有効性を示した．両者は性質の異なるタスクであるが，英語においてはどちらもまず文中からメンションに対応するスパンを抽出する必要があるという点では共通している．英語の橋渡し照応解析では共参照解析とは対照的にコーパスが十分に整備されていなかったため，学習に多くの事例を必要とするニューラルネットワークではこのスパン抽出が十分に出来ていなかった．共参照解析を同時学習して橋渡し照応解析の精度が向上したのは，共参照解析によって橋渡し照応解析にとっての良いスパン表現が得られたのが一因である．

2.3 BERT

BERT (Bidirectional Encoder Representations from Transformers) [1] は Transformer [24] をベースとした事前学習モデルであり，タスク非依存の pre-training とタスク固有の fine-tuning の 2 ステップを通して用いられるのが一般的である．図 3 に BERT の pre-training と fine-tuning の概略を示す．本節ではまず，Transformer に基づくアーキテクチャとモデルへの入力の作成方法を述べ，続いて pre-training と fine-tuning の手法について説明する．

2.3.1 Transformer

BERT のアーキテクチャは機械翻訳の分野で提案された Transformer [24] というモデルに基づく．Transformer は RNN や CNN とは異なる self-attention 機構で構成される．self-attention 機構とは，入力テキスト中の各トークンについて他のトークンとの重要度を計算し，その重み付きで他のトークンの情報を足し合わせる仕組みである．Transformer では，この操作を繰り返すことで入力テキスト中の他のトークンの情報も加味した，文脈に基づくトークン表現を得ることが可能である．self-attention は他の全てのトークンに対して計算されるた

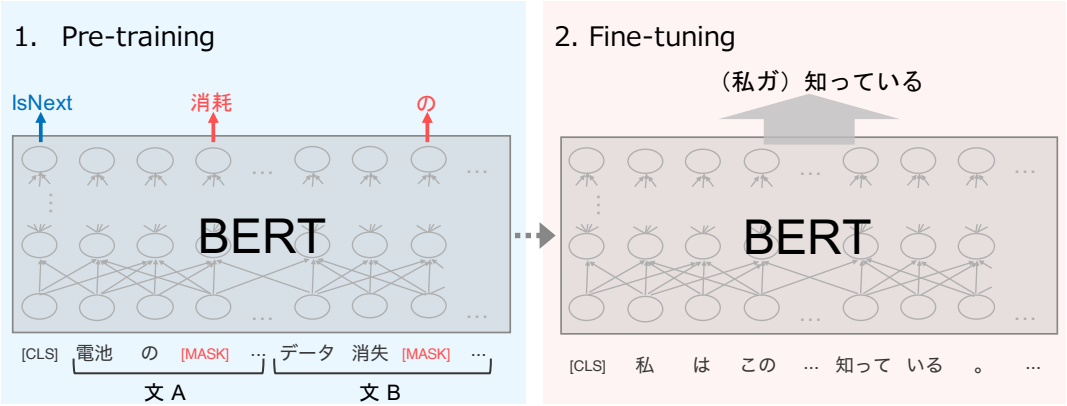

図 3: BERT の pre-training と fine-tuning．例は，fine-tuning において述語項構造解析を解く場合を示す．

め，Transformer は入力テキスト中の長距離の依存関係を捉えることが可能である．一方で，self-attention ではトークン間の計算の対称性から，RNN とは異なり各トークンの位置情報が考慮されない．したがって，Transformer では位置エンコーディングとよばれる，トークンの位置情報を表すベクトルを入力に加える．BERT においては表現力を高めるため，位置エンコーディングの代わりにトークン位置に対応する埋め込み表現 (position embedding) が使用される．

2.3.2 入力表現

はじめに，BERT へ入力されるテキストは Transformer [24] に倣い，サブワード [25] とよばれる単位に分割される．サブワードは単語よりも小さい単位であり，低頻度語が分割の対象となる．低頻度語は分割の結果，複数の高頻度サブワードとなるため，低頻度語や未知語に頑健な単語埋め込みを学習することが可能となる．BERT の入力表現を表す図 4 では，低頻度語 “playing” が “play” と “##ing” に分割されている．なお，分割前の単語列が再構成可能なよう，サブワードに分解された語のうち先頭ではないものには語の前に “##” が 付加される．

BERT ではサブワード分割の後，先頭に `[CLS]` トークン，末尾に `[SEP]` トークンが挿入される．`[SEP]` トークンは入力テキストに対する分類タスクを解く際に使用され，`[SEP]` トークンは異なる種類の 2 つの文を入力する際に使用される．最後に，このトークン列を token embedding, position embedding, segment embedding に変換し，それら 3 つのベクトルを加算して BERT モデルに入力する．token embedding は各サブワードトークンに対応する埋め込み表現である．

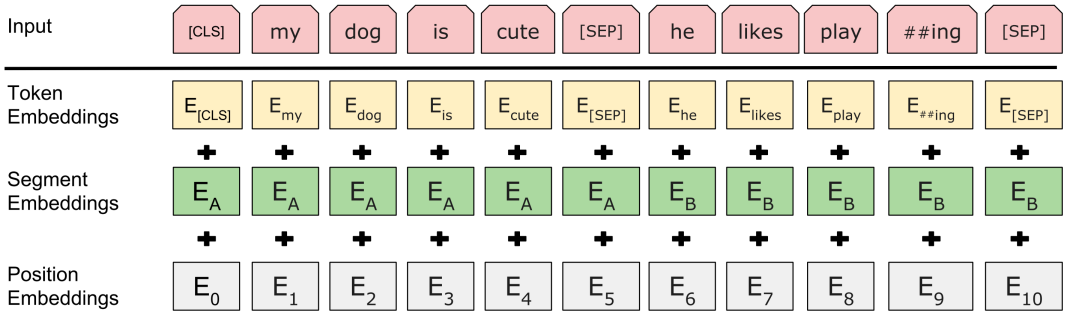

図 4: BERT の入力表現（[1] より）.

position embedding については先に述べた通りである．BERT はこの position embedding を学習する都合上，予め最大入力系列長を定めておく必要がある．segment embedding は，入力文の種類を表す埋め込み表現であり，BERT に異なる種類の 2 つの文が入力される場合，それらを判別するため 2 種類の表現が使用される．なお，この場合 2 つの文の境界を明示するため，間に `[SEP]` トークンが挿入される．

2.3.3 Pre-training

BERT の pre-training は自己教師あり学習の枠組みで行われる．自己教師あり学習は学習時の入力から機械的に教師データを作成して学習する手法である．人手で教師データを整備する場合，コストの面から大規模なデータを用意するのが困難であるが，自己教師あり学習においては容易にデータサイズを増やすことが可能である．BERT では大規模な生コーパスで pre-training を行うことで，小規模のデータからは獲得が困難な言語の汎用的な特徴を学習し，fine-tuning において高い性能を示す．

pre-training において行われるタスクには masked language model (MLM) と next sentence prediction (NSP) の 2 種類がある．MLM は入力のトークン列から一定確率で無作為にトークンをマスクし，モデルにそのマスクされたトークンを当てさせるというタスクである．マスクされたトークンの予測には周囲の単語の品詞や文の係り受け構造など様々な手がかりが必要となるため，BERT は MLM を通してそれらの特徴を学習することが期待される．NSP はモデルに異なる種類の文ペアを入力することで行われる．この際，入力データの 50%は実際にコーパス中に連続して現れる文ペアを入力し，残りの 50%はコーパスから無作為に選んだ 2 文を入力する．NSP は入力された文ペアがコーパス中で連続

していたものかどうかを判別するタスクである．NSP は fine-tuning 時に文間の関係を捉えるために導入されたが，後に性能をわずかに劣化させることが報告されている [18, 26, 27].

2.3.4 Fine-tuning

Fine-tuning は pre-training を通して学習された BERT を目的とする特定のタスクに特化させるステップである．具体的には，pre-training された BERT のモデルの上にタスク固有の最終層を加え， BERT の隠れ層のベクトルをこの最終層に入力する．この最終層の出力が上手くタスク固有の正解を予測できるように，タスクごとのデータセットで BERT モデルと最終層の重みを学習する．目的のタスクがトークンや文，あるいは文書に対する分類問題に帰着できる限り，BERT の fine-tuning は適用可能である．

第3章　日本語結束性タグ付きコーパス

本研究で扱う述語項構造解析などの解析タスクの学習には，人手で情報が付与されたコーパスの存在が不可欠である．名詞を含めた述語項構造，共参照関係，橋渡し照応関係が全て付与されたコーパスとしては以下の3つが存在する．

- 京都大学テキストコーパス[1)]
- 京都大学ウェブ文書リードコーパス[2)]
- 不満調査データセットタグ付きコーパス[3)]

それぞれのコーパスの統計を表3に示す．本章では，3.1節で統計に基づき，それぞれのコーパスの特徴を説明する．さらに，3.2節でコーパス作成時の作業者間一致度を示し，タスクの精度の上界について議論する．

3.1　各コーパスの特徴

京都大学テキストコーパス（ニュースコーパス）[28, 29]は毎日新聞の記事および社説に対してタグ付けを行ったコーパスである．コーパス全体は形態素・構文情報が付与された約40,000文から成り，このうち約10,000文に述語項構造，共参照関係，橋渡し照応関係が付与されている．

京都大学ウェブ文書リードコーパス（ウェブコーパス）[30, 31]はウェブ上に存在する種々の日本語文書に対してタグ付けを行ったコーパスである．対象となった文書は，企業・店舗ページ，ブログ・個人ページ，辞典・解説記事など多岐にわたり，それぞれページのリード文3文に対してタグが付与されている．これは，ほとんどのゼロ照応関係が3文以内に含まれているためである．表3に示すように，サイズは約15,000文でこれら全てに形態素・構文情報も付与されている．文数はニュースコーパスより多いが，1文が短いため，付与されているタグ数はニュースコーパスより少ない．サイズ以外の違いとしては，依頼，勧誘などの著者から読者へ働きかけを行うような表現の増加が挙げられる．また，そういった文体の多様化に加え，著者や読者を表す表現および，著者や読者への照応表現の増加も顕著である．

不満調査データセットタグ付きコーパス（不満コーパス）は，一般の人々か

1) https://github.com/ku-nlp/KyotoCorpus
2) https://github.com/ku-nlp/KWDLC
3) https://github.com/ku-nlp/AnnotatedFKCCorpus

表 3: 各コーパスの統計.

	ニュース	ウェブ	不満
文書数	1,909	5,124	654
文数	15,872	16,038	1,282
1 文書あたりの文数	8.3	3.1	2.0
1 文書あたりの基本句数	112.7	23.6	13.0
1 文書あたりのサブワード数	240.4	53.9	30.4
用言述語数	42,257	31,122	3,266
体言述語数	13,870	8,316	404
ガ格の項数	54,796	38,823	3,635
ヲ格の項数	21,343	15,494	1,218
ニ格の項数	13,890	11,592	1,148
ガ 2 格の項数 [2)]	1,173	2,084	432
橋渡し照応詞数	24,139	13,496	1,131
共参照メンション数	31,637	13,556	417

ら集められた不満テキストに対し，タグ付けを行ったコーパスである．不満テキストは株式会社 Insight Tech が運営するウェブサービス「不満買取センター」を通して収集された．ニュースコーパスやウェブコーパスと比べ，サイズは小さいがより口語的な表現が多いことが特徴である．また，不満テキストは一般の人々が自身が感じる不満を投稿したものであるため，必然的に著者表現や著者への照応が多い．一方で，読者表現や読者への照応は存在しない [1)]．

3.2 作業者間一致度

萩行ら [30] はウェブコーパス作成の際，述語項構造アノテーションにおける作業者間一致度を報告している．アノテーションの作業者間一致度は一般に，タスクにおける精度の上界を示す指標となる．そこで，表 4, 5 にそれぞれ用言，事態性名詞における述語項構造の一致度を示す．Shibata ら [8] の先行研究では

[1)] 1 件のみ読者への照応が存在したが，その内容からアノテーションミスと考えられる．

[2)] 例えば,「ゾウは鼻が長い。」という文では述語「長い」は複数のガ格の項を持つ．ガ 2 格はこのような場合にそれらを区別するために使用される．

表 4: 用言の述語項構造アノテーションの作業者間一致度（[30] より）.

	case	文脈ゼロ照応	外界ゼロ照応	合計
ガ	0.92	0.57	0.71	0.87
ヲ	0.93	0.66	0.46	0.88
ニ	0.91	0.44	0.49	0.78
ガ２	0.58	0.14	0.44	0.45
その他	0.72	0.27	0.36	0.67

表 5: 体言の述語項構造アノテーションの作業者間一致度（[30] より）.

	case	文脈ゼロ照応	外界ゼロ照応	合計
ガ	0.60	0.45	0.57	0.60
ヲ	0.76	0.48	0.17	0.57
ニ	0.34	0.57	0.42	0.47
ガ２	0.00	0.33	0.00	0.13
その他	0.52	0.38	0.28	0.49

ウェブコーパスにおいて用言のゼロ照応解析が，F1 スコア 58.1 で解けることが報告されているが，表 4 の文脈ゼロ照応および外界ゼロ照応の一致度はそのスコアを下回っている．また，事態性名詞に対してはさらに低い一致度となっている．これは，表の一致度が詳細なアノテーション基準を定める以前のものであるためと考えられる．萩行らは一致度の調査から得られたタグ付けが定まらない例について分析し，一意にタグ付けが可能となるよう詳細なアノテーション基準を定めた [30]．また，文脈の不足から一意に解釈ができない例など，アノテーション基準の厳密化だけでは一貫したアノテーションが不可能な例に関しては，全ての解釈についてアノテーションを行った．加えて，表 4, 5 の一致度は一人の作業者によるアノテーションを比較したものである一方，実際のアノテーション作業では一人の作業者が作業した後に別の作業者が内容の確認・修正を行う．以上から，表に示した数値は比較的低いが，最終的に作業が完了したコーパスにおいては十分な作業者間一致度が得られていると考えられる．

第4章　BERTに基づく結束性の統合的解析手法

本章では，2.3節で述べたBERT [1] を使用し，どのようにして種々の結束性解析を統合的に行うかについて説明する．まず，4.1節で手法の概要を述べ，続く4.2, 4.3節でその詳細を説明する．また，4.4, 4.5節で改良を加えた他のモデルについて説明する．

4.1　手法の概要

本研究では種々の解析を統合的に行うため，それぞれのデータを一様に混合してシステムに入力し，最後にそれぞれの結果を同時に出力する．このような解析をここでは同時解析とよぶ．

同時解析を行うためにはそれぞれのタスクを同じ枠組みで扱う必要があるが，そこで本研究では項選択モデルを使用する．項選択モデルとは，依存構造解析などでも使用されるモデルであり，テキスト中のある語を対象としてその語と係り受けなどの関係を持つ語をテキスト中から選択する枠組みである．この手続きを他の語も対象として行うことで，テキスト全体の関係を解析することができる．

本研究における解析全体の手続きは，以下の3つのステップから構成される．

1. 入力文書を基本句に分割
2. 基本句列から解析対象基本句を選択
3. 解析対象基本句に対して項選択モデルを使用し，それぞれの関係を予測

解析対象基本句の選択と項選択モデルに関しては，それぞれ4.2節と4.3節で詳説する．第1ステップに関して，本研究では基本句とよばれる単位で種々の結束性を解析する．基本句とは，核となる1つの自立語と，その前後に存在する接辞や助詞などの付属語から構成される言語の単位である [32]．本研究では，日本語形態素解析システム Juman++ [33, 34] と日本語構文解析システム KNP [35, 36, 37] を使用して文書を基本句に分割する．

4.2　解析対象基本句の選択

述語項構造は述語に対する関係，共参照は名詞に対する関係であるので，予めタスクごとに解析の対象となる基本句を選択する．基本句の選択にはKNPを利用する．KNPは基本句への分割と同時に基本句に対し様々な素性を付与す

る [38]．この素性は Juman++ によって付与された品詞などの形態素に対する素性からルールベースで決定される．

本研究では，解析対象基本句の選択に＜用言＞，＜体言＞，＜非用言格解析＞の 3 つの素性を使用する．＜用言＞，＜体言＞はそれぞれ用言と体言に，＜非用言格解析＞はサ変名詞などの，用言ではないが格を取るような名詞に対して付与される素性である．それぞれのタスクにおける選択のルールは以下の通りである．

- 用言述語項構造解析: KNP が＜用言＞を付与した基本句
- 体言述語項構造解析: KNP が＜非用言格解析＞を付与した基本句
- 橋渡し照応解析: 体言述語項構造解析の対象ではないもののうち，KNP が＜体言＞を付与した基本句
- 共参照解析: KNP が＜体言＞を付与した基本句

例えば，例 (5) の文書からは例 (6) のように各タスクについて基本句が抽出される．ここで,「用」,「体」,「橋」,「共」はそれぞれ用言述語項構造解析，体言述語項構造解析，橋渡し照応解析，共参照解析を表す．橋渡し照応解析と共参照解析では同様の基本句が選択されているが,「訪問では」が体言述語項構造解析の対象となっているため，橋渡し照応解析では選択されていない．

(5) 私はこの小説の作者を知っている。前回の訪問では彼が原稿を見せてくれた。

(6) 私は$_{橋,\ 共}$ この 小説の$_{橋,\ 共}$ 作者を$_{橋,\ 共}$ 知っている。$_{用}$ 前回の$_{橋,\ 共}$ 訪問では$_{体,\ 共}$ 彼が$_{橋,\ 共}$ 原稿を$_{橋,\ 共}$ 見せてくれた。$_{用}$

このルールを使用することで，コーパスにアノテーションされているほとんどの基本句を解析対象とすることができる．表 6 に，実際にコーパスに関係が付与されている基本句のうち，上記のルールを使用することで解析対象として選択された基本句の割合を示す．ここで，用言述語項構造解析と体言述語項構造解析は関係セットが同じであるためコーパスのアノテーションからどちらのタスクであるかを判断することはできない．そこで，KNP が ＜用言＞ タグを述語側に付与した関係を用言述語項構造解析と見做した．これは上記の用言述語項構造解析の基本句選択ルールと同一であるため，表中の用言述語項構造解析の割合は 100% になっている．体言述語項構造解析のカバレッジは比較的低い

が，用言述語項構造解析と合わせることで述語項構造アノテーションの 94.4% をカバーすることができる．

一方，上記のルールで拾えなかった基本句の例を以下に示す．

体言述語項構造解析

(7) 投資信託の約定金額は【申込み】時点では確定していません。
（ガ格：「不特定: 人」ヲ格：「信託」）

(8) ここではクレジットカードの年会費が【無料の】カードを紹介します。
（ガ格：「会費」）

橋渡し照応解析

(9) ご飯を食べた感想は、人【それぞれ】いろんな感想があると思います。
（橋渡し照応：「人」）

(10) 飲料用のリサイクル【マークの】付いた缶に限ります。
（橋渡し照応：「リサイクル」）

共参照解析

(11) 見るとするなら、もう一度本を読んでから【見たいと】思います。
（共参照：「見るとするなら、」）

(12) そんな時、買い物に行く車の中でみかけた【のが】ここのモデルルームでした。
（共参照：「ルーム」）

体言述語項構造解析に注目すると,「申し込み」や「無料の」など本来解析対象として選択されるべき基本句が選択されていないことが分かる．＜非用言格解析＞を付与する対象を拡大していくことは今後の課題である．

橋渡し照応解析に関しては,「それぞれ」など副詞にも照応詞となるものがあることが分かる．したがって，現状，橋渡し照応解析の対象は体言に限られているが，その他の基本句も対象に加えることが考えられる．例 (10) の「マーク」は KNP によって＜非用言格解析＞の素性が付与されていたため，橋渡し照応解析ではなく体言述語項構造解析の対象となっていた．名詞「マーク」は「大

表 6: コーパスに関係が付与されている基本句のうち，解析対象として選択された基本句の割合．この数値はニュースコーパスの学習データにおけるもの．

用言述語項構造解析	100.0%
体言述語項構造解析	78.1%
橋渡し照応解析	93.5%
共参照解析	93.6%

会記録をマークする」といった文脈では事態性を持ち，項を取ると考えられるが，例 (10) の文脈では事態性が失われている．

共参照解析に注目すると，例 (11) のように用言が含まれていることが分かる．コーパスでは用言に対する共参照関係もアノテーションの対象となっているが，本研究では体言に対する共参照関係のみを対象とする．また，少数ながら例 (12) のように形式名詞に対する共参照関係も見られた．

体言述語項構造解析と橋渡し照応解析の例から＜非用言格解析＞の付与，すなわち名詞の事態性の判定に課題が残されていることが分かる．これは文脈にも依存することからルールベースでは網羅しきれない可能性がある．したがって，ニューラルモデルの使用も視野に入れつつ事態性判定を改善していくことが今後の課題である．

4.3 同時解析モデル

本研究では結束性の統合的解析のため，BERT を使用し，用言述語項構造解析，体言述語項構造解析，橋渡し照応解析，共参照解析を全て項選択の枠組みで同時解析する．このモデルを Cohesion Analysis モデル (CAModel) とよぶ．

図 5 に例として CAModel を使用し，用言述語項構造解析を解く場合を示す．図は，述語「知って」のガ格の項を解析する場合を表している．CAModel は文中の他の全てのサブワードについて述語「知って」のガ格の項らしさのスコア $s_{\mathrm{NOM}}(t_i, t_j)$ を計算し，その中で最もスコアが高かったものを「知って」のガ格の項として採用する．図 5 では「知って」のガ格の解析のみを示したが，この計算をヲ格など他の格について，さらに文中の全ての述語に対してそれぞれ行う．体言述語項構造解析の場合は述語の代わりに事態性名詞を対象とすることで用言述語項構造解析と同様に解くことができる．共参照解析については，文書中

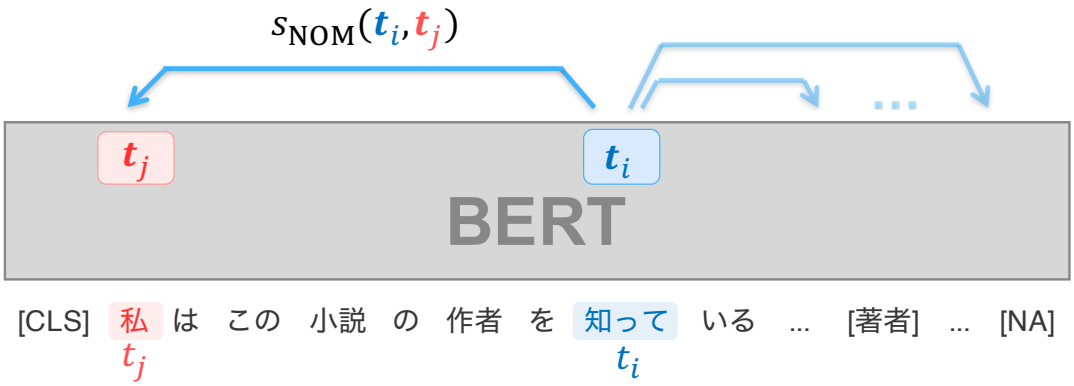

図 5: CAModel の概略．用言述語項構造解析において述語「知って」のガ格 (Nominative) の項を解析する場合を示す．

の体言に対してエンティティを共有するメンションを選択する．橋渡し照応解析については，文書中の体言に対して橋渡し照応関係にある体言を選択する．

このようにして，$s_c(t_i, t_j)$ からサブワード t_i に対して関係 c を持つ サブワード t_j を得ることができるが，4.2 節で選択したのは解析対象はサブワード単位ではなく基本句単位である．したがって，項選択モデルを使うには基本句をサブワードへ対応付ける必要がある．本研究では，基本句に含まれる自立語をその基本句を代表する形態素とし，さらにその代表形態素が複数のサブワードに分割された場合はその中の先頭のものを形態素を代表するサブワードとする．また，出力 t_j に関しては t_j を含む基本句をシステムの出力とする．

項選択モデルでは系列中から必ず一つのサブワードが選択されるが，述語や照応詞の中には項や先行詞を取らないものも存在する．また，外界照応の場合，文書中に対象は出現しない．そこで，入力系列に特殊トークンを加えることでこれらの場合に対応する．本研究では，特殊トークンとして ［著者］・［読者］・［不特定: 人］・［不特定: 物］・[NULL]・[NA] の 6 つを使用する．［著者］・［読者］・［不特定: 人］・［不特定: 物］ は外界照応解析におけるターゲットとして使用される．[NULL]・[NA] はそれぞれ，述語項構造解析おいて項が存在しない場合，共参照解析において共参照先が存在しない場合のターゲットである．

4.3.1 入力系列

同時解析モデルへの入力は以下のようにして作成される．まず，入力文書に対して文区切りと形態素分割を施した後，BPE [25] を用いてサブワードに分割する．文区切りと形態素分割に関してはコーパスに正解が付与されているので，

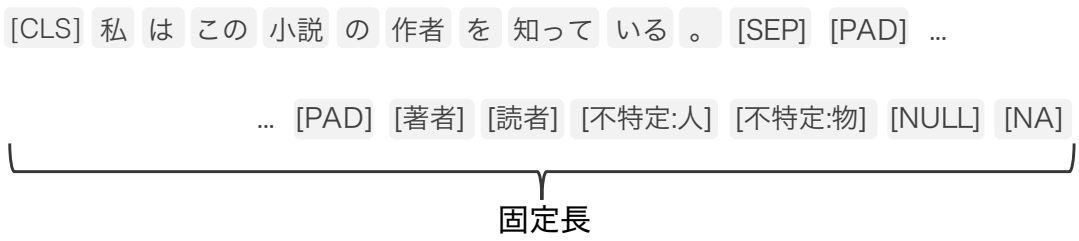

図 6: モデルへの入力系列の例

それを用いる．次に，Devlin ら [1] に倣い，入力文書の先頭と末尾にそれぞれ [CLS] トークンと [SEP] トークンを挿入する．さらに，特殊トークン［著者］・［読者］・［不特定: 人］・［不特定: 物］・[NULL]・[NA] を入力系列の末尾に挿入する．例えば,「私はこの小説の作者を知っている。」というテキストからは,図 6 に示す入力系列が作成される．ここで，[PAD] は事例ごとの系列長を揃えるのためのパディングトークンである．

図 6 に示すように，BERT への入力は固定長であることが要求される．したがって，この長さを超えるような文書に対しては文書を分割して解析する必要がある．このとき，文書を文の途中で分割することはせず，解析対象の文ができるだけ多くの先行する文を持つよう分割する．図 7 にこの分割手法を簡略化して示す．図はサブワード数がそれぞれ 20, 40, 30, 50, 20, 50 である 6 つの文を含む文書を，各文書のサブワード数が 100 以下になるよう分割する場合を表す．図の青色で示した文が解析対象の文である．対象でない文は分割後の他の文書で解析対象となるため，コンテキストとして入力には使用するが解析は行わない．

4.3.2 出力層

同時解析モデルは BERT と，BERT の出力から予測を行う出力層から構成される．出力層では，任意のサブワードペアに対し，関係 c で結ばれるかどうかを表すスコア $s_c(t_i, t_j)$ を計算する．ここで，関係 c は，用言述語項構造解析，体言述語項構造解析においては格に対応し，対象とする格と同数存在する．橋渡し照応解析，共参照解析については，用言述語項構造解析，体言述語項構造解析と異なり格ごとの解析が必要ないのでそれぞれ 1 つずつ関係が存在する．

出力層には多層パーセプトロンを使用する．具体的には，$s_c(t_i, t_j)$ は以下のように計算される．

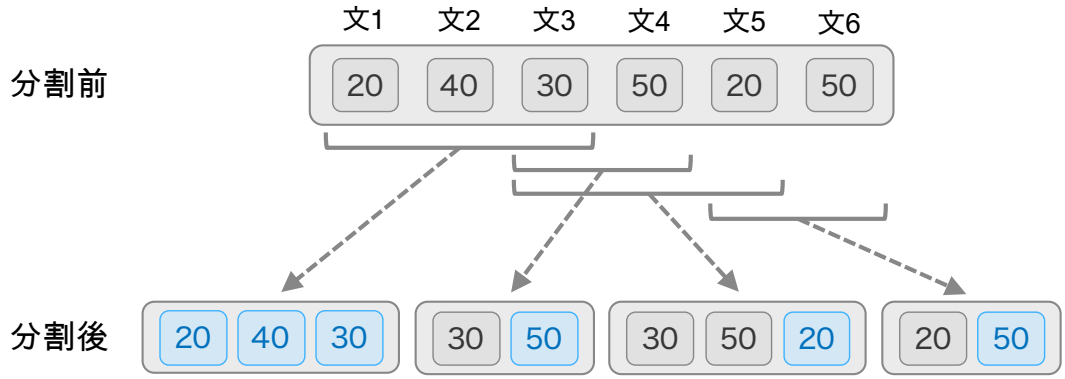

図 7: 文書の分割例．数字の書かれたセルは 1 文を表し，数字は文中に含まれるサブワードトークン数を表す．図では最大トークン数が 100 以下になるよう分割している．

$$s_c(t_j, t_i) = \boldsymbol{v}^{\mathrm{T}} \tanh(W_c \boldsymbol{t}_j + U_c \boldsymbol{t}_i), \tag{1}$$

ここで，W_c, U_c はそれぞれ関係ごとの重み行列である．$\boldsymbol{v}$ はそれぞれの関係で共通の重みである．$\boldsymbol{t}_i, \boldsymbol{t}_j$ はサブワード t_i, t_j に対応する BERT 最終層のベクトルを表す．

4.4 共参照解析を特別に扱ったモデル

本研究で扱うタスクの中でも，共参照解析は他のタスクとは性質が異なり，単純な同時解析では他のタスクの結果に寄与しないと考えられる．また，後の実験でも示されるが，共参照解析は他のタスクと同時解析を行わず，単独で解析する場合に最も高いスコアを示す．そこで共参照解析のみ特別に扱う Coreference-aware Cohesion Analysis モデル (CorefCAModel) を提案する．

モデルの概略を図 8 に示す．CorefCAModel による予測は 2 つのステップからなる．まず第 1 ステップでは，共参照解析は単独で解くほうが良いことを踏まえ，入力に対して共参照解析のみを行う．続く第 2 ステップでは全てのタスクを解析する．このとき，第 1 ステップの結果を素性として利用する．これは，共参照関係の情報が他のタスクにとって有用であると期待されるためである．例えば以下の例で述語「心配している」のヲ格の項を解析することを考える．ヲ格の項の候補としては,「太郎」や「次郎」が考えられるが,「太郎」と「彼」が共参照関係にあることを踏まえると「太郎」ではないことが明確になる．

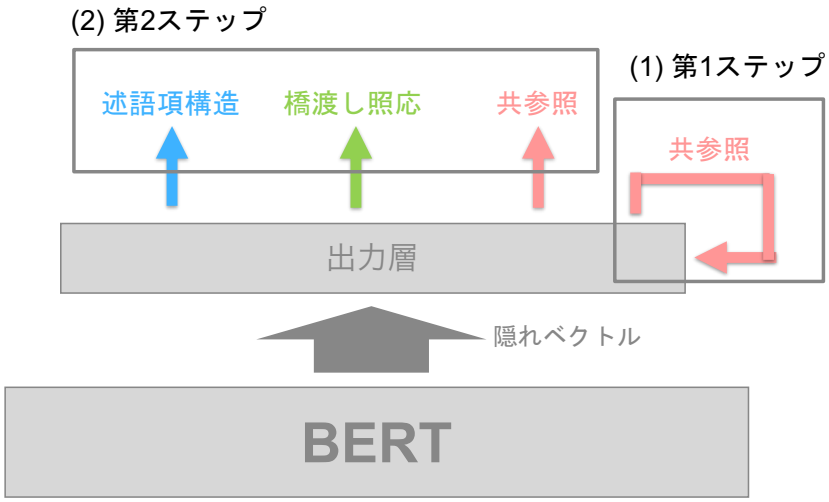

図 8: CorefCAModel の概略.

(13)　太郎は次郎と連絡がつかないらしい。彼はとても心配している。

具体的に，CorefCAModel によるスコア $s_c(t_i, t_j)$ は以下の式で計算される．

$$s_c(t_j, t_i) = \boldsymbol{v}^{\mathrm{T}} \tanh(\boldsymbol{h}_{i,j,c}) \tag{2}$$

$$\boldsymbol{h}_{i,j,c} = W_c \boldsymbol{t}_j + U_c \boldsymbol{t}_i + \boldsymbol{h}_j^{coref} \tag{3}$$

$$\boldsymbol{h}_j^{coref} = \sum_k P_{coref}(t_k|t_j) V \boldsymbol{t}_k, \tag{4}$$

ここで，$\boldsymbol{v}, V$ は全ての関係で共通の重み行列，W_c, U_c は関係ごとの重み行列である．$\boldsymbol{t}_i, \boldsymbol{t}_j, \boldsymbol{t}_k$ は入力トークン t_i, t_j, t_k に対応する BERT 最終層のベクトルである．$P_{coref}(t_k|t_j)$ は第 1 ステップの結果であり，トークン t_j がトークン t_k と共参照関係にある確率を表す．これは，CAModel において共参照解析のみ解いた場合に相当し，以下の式で計算される．

$$P_{coref}(t_k|t_j) = \frac{\exp(s'_{coref}(t_k, t_j))}{\sum_l \exp(s'_{coref}(t_l, t_j))} \tag{5}$$

$$s'_{coref}(t_k, t_j) = \boldsymbol{v}^{\mathrm{T}} \tanh(W_{coref} \boldsymbol{t}_k + U_{coref} \boldsymbol{t}_j). \tag{6}$$

4.5 係り受けを考慮したモデル

詳細は 5.3.2 節で述べるが，格解析においては提案手法である CAModel が既存手法を下回ることが分かった．これは，既存手法が選択選好や係り受けなど，多くの素性を使用しているためと考えられる．特に，述語と項の間の係り受けの情報は格解析において非常に有効であることが知られている [12]．そこで，格解析の精度向上のため，本研究でも追加の素性として係り受けの使用を提案する．この係り受けを考慮したモデルを Dependency-aware Cohesion Analysis モデル (DepCAModel) とよぶ．

モデルの概略を図 9 に示す．図は，述語「知って」のガ格の項が「私」であるスコアを計算する場合を示している．CAModel との違いは，スコアの計算時に係り受け関係の情報を与えるための埋め込みベクトル $\boldsymbol{d}$ が追加されていることである．埋め込みベクトル $\boldsymbol{d}$ は $\boldsymbol{d_0}, \boldsymbol{d_1}, \boldsymbol{d_2}$ の 3 種類用意されており，それぞれ項候補から述語への係り受け，述語から項候補への係り受け，係り受けなしに対応する．図では，項候補「私」が述語「知って」に係っているため，$\boldsymbol{d_0}$ が s_{NOM} の計算に使用される．

一般の関係 c について考えると，DepCAModel によるスコア $s_c(t_j|t_i)$ は以下の式で計算される．

$$s_c(t_j, t_i) = \boldsymbol{v}^{\mathrm{T}} \tanh(\boldsymbol{h}_{i,j,c}) \tag{7}$$

$$\boldsymbol{h}_{i,j,c} = W_c \boldsymbol{t}_j + U_c \boldsymbol{t}_i + \boldsymbol{d}_{i,j} \tag{8}$$

$$\boldsymbol{d}_{i,j} = \mathrm{Emb}(\mathrm{dep}(t_j, t_i)), \tag{9}$$

ここで，CAModel と同様，$\boldsymbol{v}$ は全ての関係で共通の重み行列，W_c, U_c は関係ごとの重み行列である．$\boldsymbol{t}_i, \boldsymbol{t}_j, \boldsymbol{t}_k$ は入力トークン t_i, t_j, t_k に対応する BERT 最終層のベクトルである．$\mathrm{dep}(t_j, t_i)$ はトークン t_i, t_j の間の係り受け関係を返す 3 値関数で，Emb はそれを埋め込みベクトルに写像する関数である．

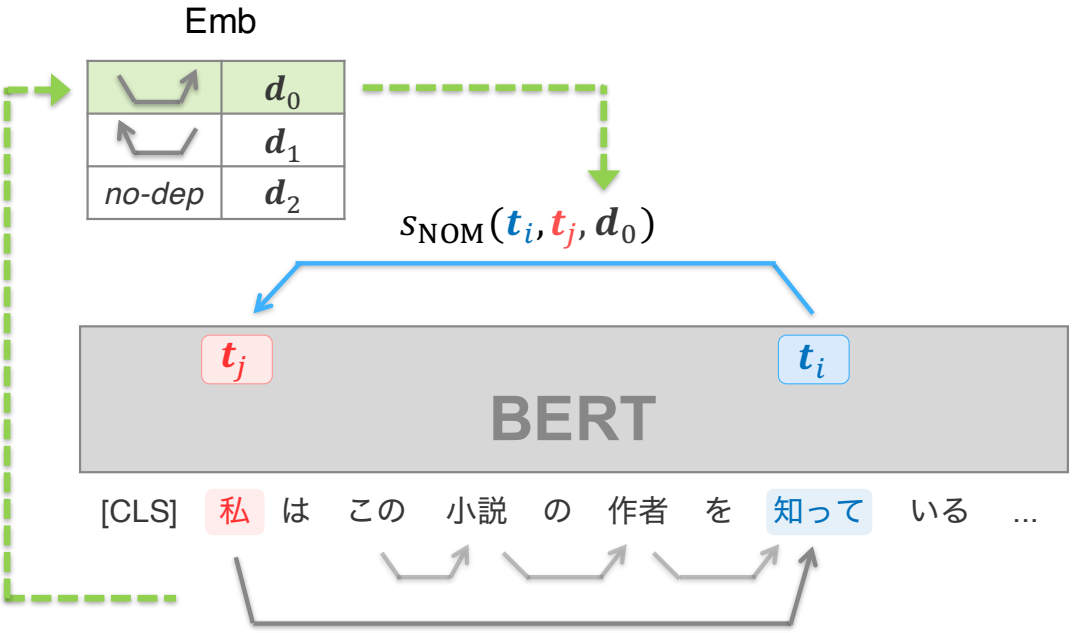

図 9: DepCAModel の概略.

表 7: 各コーパスのデータ分割の統計（文数）.

	学習	開発	テスト
ウェブ	12,258	1,585	2,195
ニュース	12,944	1,145	1,783
不満	885	195	202

第5章　実験

本章では 第 4 章で説明した手法に基づく実験について述べる．5.1, 5.2 節で実験設定および評価手法を整理し，5.3, 5.4, 5.5 節で提案手法の評価を行う．また，5.6 節で実験結果について議論する．

5.1　実験設定

実験にはニュースコーパス，ウェブコーパス，不満コーパスの 3 種類のコーパスを使用した．それぞれ，表 7 の通りに学習，開発，テストデータに分割した．学習は全てのコーパスを混合したものを使用して行い，評価はそれぞれのコーパスについて行った [1)]．

実験では計算資源の都合上，BERT への入力トークン数を 128 とした．表 3 に示すようにニュースコーパスは 1 文書あたり平均 240.4 のサブワードトークンを含んでおり，多くの文書が 128 の長さを超えてしまうため，1 文書を分割して系列長を 128 以下にした後，学習および評価を行った．ウェブコーパスに関してはいずれの文書も系列長は 128 よりも短かった．不満コーパスに関しては 1 例のみテストデータに 128 を超える系列長のものが存在したが，分割はせず単に評価セットから除外した．

ニュースコーパスの分割数の分布は図 10 のようになった．分割されなかった文書は 608 件あり，全体のおよそ 32%だった．ニュースコーパスのテストデータにおいて，分割により正解候補から外れた基本句の割合をタスクごとに表 8 に示す．表から，述語項構造や橋渡し照応においては，多くの項・先行詞が 128 トークンの中に収まっていることが分かる．一方で，共参照は文書中の広範に

[1)] 予備実験において，各コーパスを単独で学習するよりも，混合して学習した場合に良好な結果が得られることを確認した．

表 8: ニュースコーパスの分割の結果，正解候補から外れた基本句の割合（テストデータ）

関係	割合
用言述語項構造	3.0%
体言述語項構造	4.7%
橋渡し照応	5.5%
共参照	28.6%

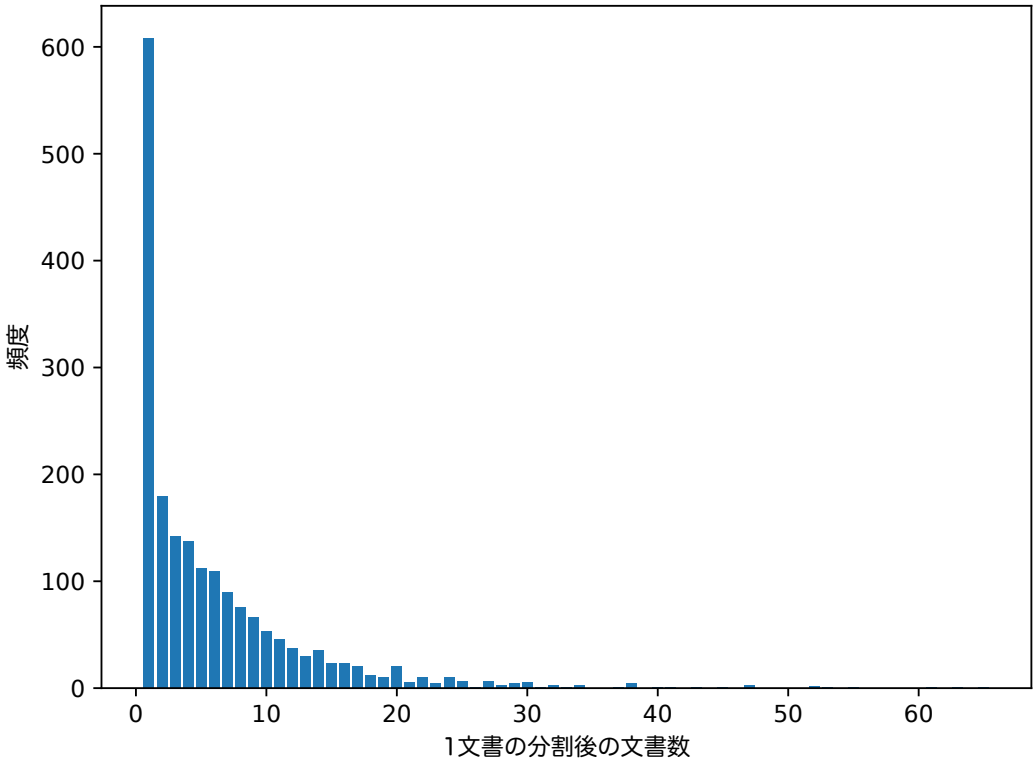

図 10: ニュースコーパス分割数のヒストグラム．

わたって存在しており，約 30%が正解候補から外れてしまった．これは共参照解析における再現率の上界が約 70%であることを意味しており，共参照解析のスコア向上には，今後の課題としてより長い文書を扱えるようになることが不可欠である．

実験には，第 4 章で述べた CAModel, CorefCAModel, DepCAModel を使用した．CAModel に関しては様々なタスクの組み合わせで学習を行った．いずれも

学習にはクロスエントロピーロスを使用し，Devlin ら [1] に倣い，最大 4 エポック fine-tuning した．その他のハイパーパラメータを表 9 に示す．

表 9: 実験に使用したハイパーパラメータ．

パラメータ名	値
オプティマイザ	AdamW
学習率	5×10^{-5}
AdamW の ϵ	1×10^{-8}
重み減衰	0.01
ドロップアウト率 (BERT 層)	0.1
ドロップアウト率 (出力層)	0.0
学習率スケジューラ	linear_schedule_with_warmup[1)]
スケジューラウォームアップ率	0.1
バッチサイズ	8
隠れ層の次元数	768

BERT モデルには，NICT BERT 日本語 Pre-trained モデル BPE あり [2)] を使用した．これは，日本語 Wikipedia 全文に対して形態素分割およびサブワード分割を行い，およそ 100 万ステップ pre-training したものである．

用言述語項構造解析，体言述語項構造解析については，ガ格・ヲ格・ニ格・ガ２格の４つの格を解析対象とした．Shibata ら [8] は外界照応の照応先として［著者］・［読者］・［不特定: 人］を扱ったが，本研究では加えて［不特定: 物］も扱った．これは，特に不満コーパスで［不特定: 物］が頻繁に出現するためである．

共参照解析については対象とするメンションより前に存在する基本句のみを予測の対象とし，後方に存在するものは学習および評価から除外した．述語項構造解析におけるゼロ照応解析および橋渡し照応解析は後方の基本句も対象とするが，後方の基本句が文をまたぐ場合は現象として稀であるため対象とせず，学習および評価から除外した．

[1)] Transformers [39] の実装を使用．

[2)] https://alaginrc.nict.go.jp/nict-bert/index.html

表 10: 正解データの作例.

	ガ格	ヲ格	共参照
太郎は			
パンを			
買って	太郎	パン	
彼の			太郎
家で			
食べた。	太郎	パン	

表 11: システム出力の作例.

	ガ格	ヲ格	共参照
太郎は			
パンを			太郎
買って	太郎		
彼の			パン
家で			
食べた。	彼	パン	

表 12: 表 10,11 のデータにおける F1 スコア.

	ガ格	ヲ格	共参照
適合率	2/2	1/1	0/2
再現率	2/2	1/2	0/1
F1 スコア	1.0	0.67	0.0

5.2 評価手法

本研究では扱う全てのタスクについて Shibata ら [8] に倣い，評価指標として F1 スコアを使用した．例として，表 10, 11 に示すような正解データとシステム出力が得られた場合の述語項構造解析と共参照解析の評価を説明する．ガ格に注目すると，述語「買って」に対してはシステムは正しく「太郎」を予測できている．一方，述語「食べた。」に対してはシステムは「太郎」ではなく「彼」と出力している．しかし，表 10 の通り正解データにおいて「太郎」と「彼」は共参照関係にあるので,「食べた」のガ格の項は「彼」でも正解である．したがって，表 12 に示すようにガ格における適合率と再現率はいずれも 1.0 であり，F1 スコアは 1.0 となる．ヲ格や共参照についても同様に評価を行う．

述語項構造解析における格解析かゼロ照応解析かの判断には，正解データに付与されている係り受け情報を使用した．全ての格に渡って F1 スコアを計算する場合は，それぞれの格における結果のマイクロ平均を取った．

5.3 CAModel

5.3.1 概要

BERT を使用した項選択モデルの評価を行うため，CAModel を使用してまず既存手法との比較を行う．さらに，同時解析の効果を検証するため，様々なタスクの組み合わせで CAModel を学習し，評価を行う．

5.3.2 結果

表 13, 14 にウェブコーパスとニュースコーパスそれぞれにおける既存研究との比較を示す．以後，表において用言述語項構造解析，体言述語項構造解析，橋渡し照応解析，共参照解析をそれぞれ VPA (Verbal Predicate-argument structure Analysis), NPA (Nominal Predicate-argument structure Analysis), BAR (Bridging Anaphora Resolution), CR (Coreference Resolution) と表記する．表中の zero, case はそれぞれゼロ照応解析，格解析を表す．zero+case はそれらのマイクロ平均，zero+case+overt は overt も加えたマイクロ平均である．Shibata ら [8] と Kurita ら [9] の研究では，zero+case のスコアが報告されていなかったため，適合率と再現率が等しいという仮定のもと算出した F1 スコアの概算値を示す．

表から，いずれのコーパスにおいても CAModel はゼロ照応解析と共参照解析の，特に難しいとされる 2 つのタスクで大きくスコアが向上していることが分かる．格解析に関しては既存研究を下回っているが，述語項構造解析全体のスコアである zero+case を比較すると Shibata&Kurohashi18 を大きく上回っている．zero+case について Kurita+18 と比較すると，スコアの差はわずかである．しかし， Kurita らはゼロ照応解析の中でも特に難しい文間ゼロ照応解析を扱っておらず，外界照応についても著者と読者しか対象としていないことを踏まえると実際の性能には大きな差があると考えられる．

表 15 に CAModel を異なるタスクの組み合わせで実験した場合の用言述語項構造解析，体言述語項構造解析，橋渡し照応解析，共参照解析のスコアをそれぞれ示す．表の左端にそれぞれ学習時に同時解析したタスクを示す．スコアの括弧内に異なる初期値で 10 回実験を行った場合の t 分布に基づく 95% 信頼区間を示す．また，図 11 に各タスクを単体で行った場合と，別のタスクを加えて学習を行った場合を比較したときの t 検定における p 値を示す．この値は 3 つのコーパスにおけるスコアのミクロ平均から計算される．p 値は 0 から 1 の間の

表 13: ウェブコーパスにおける CAModel の既存研究との比較.

	VPA				BAR	CR
	zero	case	zero+case	zero+case +overt		
Sasano&Kurohashi09 [15][1)]	-	-	-	-	42.7	-
Shibata&Kurohashi18 [8]	58.1	89.2	70.6	-	-	68.5
Kurita+18 [9][2)]	58.4	**92.0**	75.7	-	-	-
CAModel (ours)	**70.0**	87.0	**76.5**	82.9	**66.2**	**76.5**

表 14: ニュースコーパスにおける CAModel の既存研究との比較.

	VPA				CR
	zero	case	zero+case	zero+case +overt	
Shibata&Kurohashi18 [8]	35.6	**89.5**	58.4	-	54.1
CAModel (ours)	**55.7**	84.8	**68.7**	79.7	**57.6**

値だが，便宜上スコアが低下した場合は負の値として表示している．

用言述語項構造解析と体言述語項構造解析は関係セットが同じであることもあり，非常に類似したタスクと考えられる．一般に，類似したタスクは同時解析によって互いに利得があることが知られている．実際，表 15 において用言述語項構造解析のスコアは体言述語項構造解析との同時解析により向上している [3)]．逆に，体言述語項構造解析も単体で学習するより用言述語項構造解析を加えたほうが高い性能を示すことが分かる．この傾向は図 11 においても確認できる．

[1)] 本研究と評価に使用したコーパスは同じだが，評価データのサイズが異なるため，厳密には比較ができない．

[2)] Kurita らは本研究はと異なり文間のゼロ照応解析を扱っていないため，zero および zero+case のスコアは厳密には比較ができない．

[3)] 不満コーパスにおいては低下しているが，表 3 から不満コーパスの文数は他のコーパスの 10 分の 1 程度である．3 コーパスを合わせて比較するとやはり体言述語項構造解析は用言述語項構造解析のスコアを向上させる．

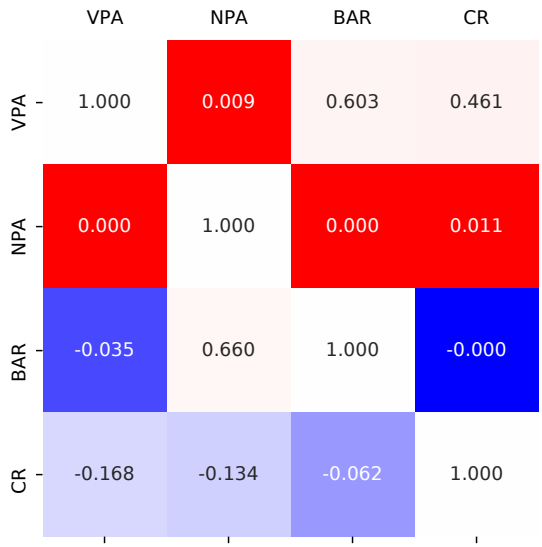

図 11: それぞれのタスクを単体で解いた場合と他のタスクと同時に解いた場合のスコアの変化．縦軸がベースとなるタスクで，横軸に示すタスクと同時解析した場合と比べたときの t 検定における p 値をセル内に示す．同時解析によるスコアの変化量で色付けしている（赤が増加，青が減少，変化量が統計的に大きいほど濃い色）．

表 15: 学習タスクの組み合わせを変えた場合の各タスクの F1 スコア.

タスク	追加学習タスク	ニュース	ウェブ	不満
VPA	*only*	68.46 (±0.34)	75.73 (±0.20)	67.85 (±0.70)
	+ CR	68.40 (±0.26)	76.02 (±0.22)	67.38 (±0.86)
	+ NPA	69.06 (±0.28)	76.20 (±0.30)	66.57 (±0.87)
	+ BAR	68.97 (±0.24)	75.42 (±0.23)	67.59 (±0.49)
	+ NPA + BAR	**69.27** (±0.18)	76.15 (±0.24)	67.23 (±0.67)
	+ NPA + BAR + CR	68.70 (±0.29)	**76.45** (±0.24)	**68.57** (±0.56)
NPA	*only*	52.72 (±0.50)	57.56 (±0.71)	**64.88** (±2.08)
	+ CR	52.90 (±0.46)	59.32 (±0.37)	61.75 (±1.70)
	+ VPA	55.14 (±0.41)	60.72 (±0.40)	60.37 (±1.93)
	+ BAR	54.57 (±0.29)	58.88 (±0.48)	61.86 (±2.20)
	+ VPA + BAR	**56.06** (±0.36)	**60.78** (±0.58)	63.17 (±1.32)
	+ VPA + BAR + CR	55.13 (±0.69)	60.35 (±0.36)	59.03 (±1.19)
BAR	*only*	**61.30** (±0.54)	67.02 (±0.30)	68.26 (±0.89)
	+ CR	59.71 (±0.26)	66.70 (±0.37)	68.56 (±1.16)
	+ VPA	60.90 (±0.37)	66.63 (±0.32)	67.72 (±1.02)
	+ NPA	61.29 (±0.60)	**67.21** (±0.29)	68.80 (±1.13)
	+ VPA + NPA	61.18 (±0.44)	67.12 (±0.39)	67.48 (±1.46)
	+ VPA + NPA + CR	60.06 (±0.39)	66.18 (±0.25)	**70.29** (±0.98)
CR	*only*	**59.35** (±0.62)	**77.73** (±0.46)	77.97 (±1.73)
	+ VPA	59.12 (±0.33)	77.19 (±0.43)	77.46 (±1.85)
	+ NPA	58.94 (±0.30)	77.25 (±0.34)	**79.29** (±1.03)
	+ BAR	58.82 (±0.38)	77.16 (±0.25)	79.27 (±1.49)
	+ VPA + NPA	58.83 (±0.38)	76.78 (±0.33)	76.52 (±1.88)
	+ VPA + NPA + BAR	57.62 (±0.48)	76.52 (±0.35)	77.83 (±0.53)

橋渡し照応解析に関して，他のタスクと同時解析することでスコアが有意に向上することはなかったが，共参照解析と同時解析した場合にスコアが低下することが分かった．これは，橋渡し照応解析と共参照解析はどちらも名詞間の関係解析だが，共参照解析が同一のものを見つけるタスクであるという点で橋渡し照応解析とは大きく性質が異なるためだと考えられる．

橋渡し照応解析にとって利得となるタスクはなかったが，一方で橋渡し照応解析は体言述語項構造解析のスコアに寄与している．図 11 からも橋渡し照応解析を解くことで体言述語項構造解析が有意に向上することが分かる．これは橋渡し照応解析と体言述語項構造解析がいずれも名詞間の関係解析タスクかつ，異なるエンティティ間の関係についてのタスクであるためと考えられる．

次に，共参照解析に注目する．共参照解析に対しては単体で解析する場合が最も良く，用言述語項構造解析，体言述語項構造解析，橋渡し照応解析いずれのタスクもスコアを低下させるという結果になった．また，共参照解析は橋渡し照応解析のスコアを低下させた．体言述語項構造解析のスコアには貢献しているが，これは体言述語項構造解析のデータ量が限られているためと考えられる．実際，表 15 において体言述語項構造解析に加え，用言述語項構造解析と橋渡し照応解析も解く条件下では共参照解析は体言述語項構造解析のスコアを低下させる．この結果からも共参照解析が他のタスクと比べ異種であるということが確認できる．

5.4 CorefCAModel

5.4.1 概要

同時解析の結果からも共参照解析が他のタスクと性質が異なることが確認された．そこで，その事実を考慮した CorefCAModel を作成し，実験を通してその性能を評価する．

CorefCAModel は学習初期，十分な精度で共参照解析が行えないため，第 2 ステップでのモデルの学習が正しく行われないおそれがある．そのため，Scheduled Sampling [40] の考えに基づき，学習時は第 1 ステップの出力に正解データを混ぜ，学習の進行に伴ってその割合を減少させた．

5.4.2 結果

表 16 に CorefCAModel の結果を示す．CorefCAModel を使用することで共参照解析には効果があったものの，他のタスクでは有意差が見られなかった．こ

表 16: CorefCAModel の結果.

タスク		ニュース	ウェブ	不満
VPA	CAModel	68.70 (±0.29)	**76.45** (±0.24)	**68.57** (±0.56)
	CorefCAModel	**69.05** (±0.25)	76.42 (±0.10)	68.02 (±0.88)
	p 値	0.054	0.799	0.250
NPA	CAModel	55.13 (±0.69)	**60.35** (±0.36)	59.03 (±1.19)
	CorefCAModel	**55.23** (±0.38)	60.19 (±0.62)	**59.17** (±1.57)
	p 値	0.770	0.602	0.879
BAR	CAModel	60.06 (±0.39)	**66.18** (±0.25)	**70.29** (±0.98)
	CorefCAModel	**60.12** (±0.33)	65.95 (±0.24)	69.91 (±1.12)
	p 値	0.788	0.160	0.571
CR	CAModel	57.62 (±0.48)	76.52 (±0.35)	77.83 (±0.53)
	CorefCAModel	**58.13** (±0.22)	**76.65** (±0.20)	**78.26** (±1.04)
	p 値	**0.046**	0.466	0.422

れは，共参照解析の精度がそれほど高くなく，他のタスクにとって利得となるほどの情報が与えられなかったためと考えられる．

5.5 DepCAModel

5.5.1 概要

表 13,14 から，CAModel を使用することで既存手法と比べてゼロ照応解析のスコアが大きく向上することが分かった．しかしその一方で，格解析のスコアは両コーパスにおいて低下している．これは，既存手法が選択選好や係り受けなどの素性を使用している一方，本研究の手法ではそれらを一切使用していないためと考えられる．そこで，格解析の精度向上のため，本研究でも追加の素性として係り受けを使用した DepCAModel を提案し，実験を通して素性の効果を検証する．

表 17: DepCAModel の結果

タスク		ニュース	ウェブ	不満
VPA	CAModel	**68.70** (±0.29)	**76.45** (±0.24)	**68.57** (±0.56)
	DepCAModel	68.54 (±0.24)	75.88 (±0.19)	66.64 (±0.84)
	p 値	0.357	**0.001**	**0.001**
NPA	CAModel	**55.13** (±0.69)	60.35 (±0.36)	59.03 (±1.19)
	DepCAModel	55.07 (±0.32)	**61.24** (±0.39)	**63.13** (±1.82)
	p 値	0.852	**0.001**	**0.001**
BAR	CAModel	60.06 (±0.39)	**66.18** (±0.25)	**70.29** (±0.98)
	DepCAModel	**60.33** (±0.38)	66.06 (±0.24)	70.17 (±1.80)
	p 値	0.280	0.467	0.895
CR	CAModel	57.62 (±0.48)	**76.52** (±0.35)	77.83 (±0.53)
	DepCAModel	**58.23** (±0.33)	76.47 (±0.33)	**78.13** (±1.23)
	p 値	**0.029**	0.832	0.627

表 18: DepCAModel の結果（正解の係り受け情報を使用）

タスク		ニュース	ウェブ	不満
VPA	CAModel	68.70 (±0.29)	76.45 (±0.24)	**68.57** (±0.56)
	DepCAModel	**69.59** (±0.24)	**76.59** (±0.18)	67.31 (±0.86)
	p 値	**0.000**	0.299	**0.014**
NPA	CAModel	55.13 (±0.69)	60.35 (±0.36)	59.03 (±1.19)
	DepCAModel	**55.15** (±0.33)	**61.28** (±0.37)	**63.13** (±1.82)
	p 値	0.948	**0.001**	**0.001**
BAR	CAModel	60.06 (±0.39)	66.18 (±0.25)	**70.29** (±0.98)
	DepCAModel	**60.47** (±0.37)	**66.41** (±0.25)	70.17 (±1.80)
	p 値	0.102	0.159	0.895
CR	CAModel	57.62 (±0.48)	76.52 (±0.35)	77.83 (±0.53)
	DepCAModel	**58.23** (±0.33)	**76.57** (±0.33)	**78.13** (±1.23)
	p 値	**0.031**	0.803	0.627

実験の際，係り受けの情報は BERTKNP [5] による自動解析結果を使用した．先行研究 [8, 9] は係り受けの情報としてコーパスに付与されている正解データを使用していたため，自動解析だけでなく正解の係り受けを与えるという条件でも実験を行った．なお，コーパスに付与されている係り受け関係は基本句単位であるため，サブワードトークン間の係り受けとしては各トークンを内包する基本句同士の係り受けを使用した．

5.5.2 結果

表 17 に DepCAModel の結果を示す．係り受けの素性を加えることでウェブコーパスおよび不満コーパスで体言述語項構造解析のスコアが有意に向上した．しかし，用言述語項構造解析においては逆にスコアが有意に低下する結果となった．また，共参照解析においても部分的に係り受けの素性がスコアに寄与することが分かった．

表 18 に自動解析から得られた係り受けではなく正解の係り受けを使用した場合の DepCAModel の結果を示す．体言述語項構造解析や共参照解析においては基本句同士が係り受けの関係にあることが稀であるため，いずれのタスクでも表 17 とおおよそ同じスコアとなっている．一方，用言述語項構造解析や橋渡し照応解析では正解の係り受けを使用することで効果が見られ，ニュースコーパスでは CAModel のスコアを有意に上回っている．不満コーパスにおいては依然，係り受けの素性が用言述語項構造解析のスコアを低下させる結果となった．

表 19 に DepCAModel の既存研究との比較を示す．(oracle) は正解の係り受けを使用したことを表す．正解の係り受けを使用する設定では，期待通り格解析のスコアが CAModel に比べて改善したものの，既存研究のスコアには及ばなかった．

5.6 議論

結果から，用言述語項構造解析，体言述語項構造解析，橋渡し照応解析は同時解析においておおむね相互に利得があるが，共参照解析は同時解析に寄与しないことが分かった．共参照解析は橋渡し照応解析のスコアを低下させ，さらに他のタスクと同時解析することの利得もない．用言述語項構造解析，体言述語項構造解析，橋渡し照応解析はエンティティ，あるいは述語間の関係を解析するタスクであるのに対し，共参照解析は同一のエンティティを発見するタスクである．実験の結果は共参照解析が他のタスクと比べ性質の異なるものであ

表 19: DepCAModel の用言述語項構造解析における既存研究との比較.

	ニュースコーパス			ウェブコーパス		
	zero	case	zero+case	zero	case	zero+case
S&K18 [8]	35.6	**89.5**	58.4	58.1	89.2	70.6
Kurita+18 [9]	-	-	-	58.4	**92.0**	75.7
CAModel	**55.7**	84.8	68.7	**70.0**	87.0	76.5
DepCAModel	55.6	83.7	68.5	69.6	85.6	75.9
DepCAModel (oracle)	54.7	86.5	**69.6**	69.4	87.7	**76.6**

るというこの事実を裏付ける.

したがって，共参照解析は用言述語項構造解析，体言述語項構造解析，橋渡し照応解析とは別に扱うのが両者にとって好ましい．本研究では共参照解析のみを先に解析し，素性として使用する CorefCAModel を提案した．共参照解析では CorefCAModel によって単純に同時解析を行う場合よりも良好な結果が得られることを確認した．しかし，他のタスクでは CorefCAModel は CAModel と有意差が見られなかった．性質が異なるとはいえ，共参照関係の情報は他のタスクの助けになると考えられるため，共参照関係の情報をより効果的に使うことのできるモデルの作成が今後の課題である.

次に，既存手法と比べたとき唯一スコアが低下した格解析を改善するため，係り受けの素性を加えた DepCAModel を提案した．実験の結果，自動解析によって得られた係り受けを素性として加えることで，体言述語項構造解析のスコアが向上した．用言述語項構造解析のスコアは逆に低下したが，これは学習データ量の差と解析誤りが原因と考えられる．学習データ量の少ない体言述語項構造解析では解析誤りを含んでいたとしても係り受けの情報が解析の有効な手がかりとなった．一方，学習データ量の豊富な用言述語項構造解析では，係り受け解析から得られる情報を部分的に学習可能なため解析誤りを含む係り受けの情報が有効な手がかりとならず，むしろスコアの低下に繋がった.

自動解析により得られた係り受けの情報では用言述語項構造解析に効果がなかったが，正解の係り受けの情報では効果が見られた．ニュースコーパスにおいては有意にスコアが向上し，ウェブコーパスにおいてもわずかながら改善が

見られた．一方，不満コーパスでは用言述語項構造解析のスコアが低下したが，これは表 3 の通り不満コーパスは 1 文が短く，係り受けを与えることで得られる情報が少なかったためと考えられる．対照的に，1 文が長いニュースコーパスでは係り受けを与えることによる利得が大きく，用言述語項構造解析のみならず共参照解析でもスコアが有意に向上した．

実験に使用した BERT に関しては事前学習を通して中間層で係り受け関係を学習しているという報告がある [41]．しかし，DepCAModel を使用した実験を通して，BRET に明示的に係り受けなどの素性を加えることが依然効果的であることが示された．したがって，DepCAModel の格解析のスコアは既存手法にわずかに及ばなかったが，選択選好などの素性をさらに加えることでこの差を埋められる可能性がある．

続いて，システムが実際に解析した事例を分析する．以下に，下線部の述語について CAModel は解析に失敗したが，DepCAModel が解析に成功した例を載せる．

(14) しかし、ピーターは既に死んでいた。ショックを受けるサイモンに、妻アナはサイモンがピーターを殺したと告げる。
正解：（ガ格）サイモン
CAModel：（ガ格）アナ
DepCAModel：（ガ格）サイモン

(15) お越しいただいた方々、ありがとうございました。
正解：（ニ格）方々
CAModel：（ニ格）NULL
DepCAModel：（ニ格）方々

例 (14) では，述語「受ける」はサイモンに係っている．したがって,「受ける」と「サイモン」は格関係にあるが，DepCAModel は素性からその情報を獲得し，ガ格の解析に成功したと考えられる．例 (15) についても同じく,「方々」と「ありがとうございました。」の係り受け関係が考慮された結果と考えられる．

次に，全ての関係について解析結果を分析する．CAModel が解析したウェブコーパスの例を図 12 に載せる．図はスペースの都合上，文書を一部省略しており，元文書は以下である．

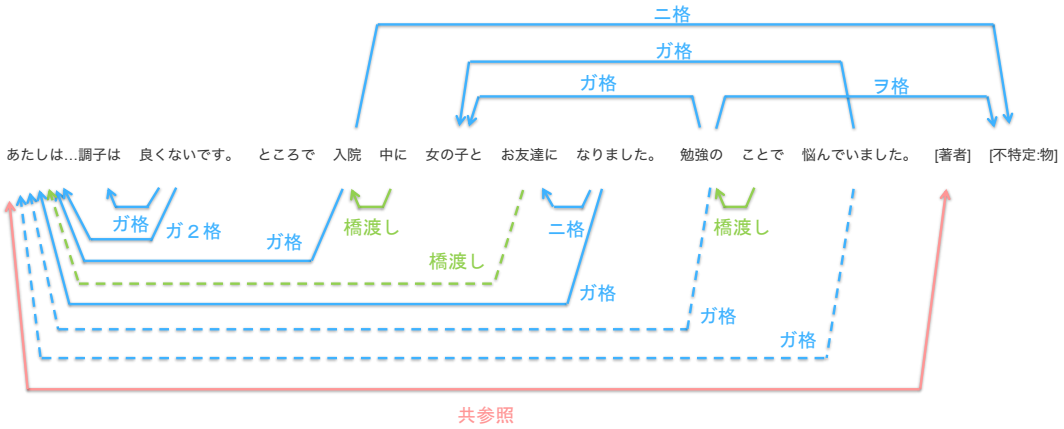

図 12: ウェブコーパスの解析例．下側の矢印はシステムの予測を表す．破線はシステムが予測を誤った例であり，上側にシステムが正しく予測できなかった正解例を示す．

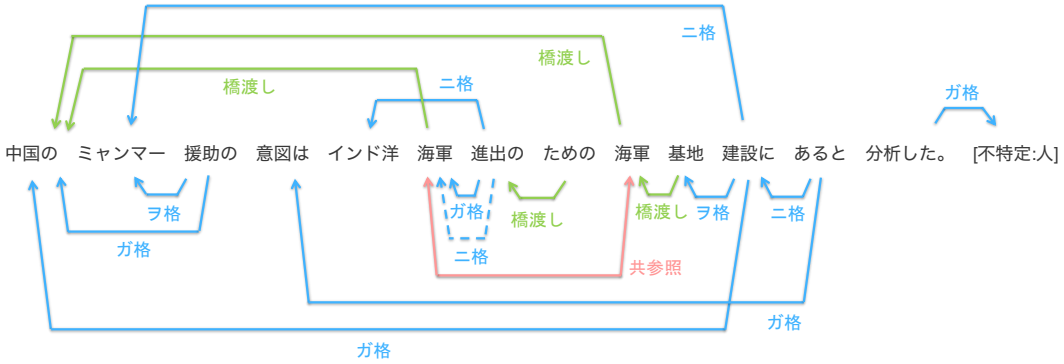

図 13: ニュースコーパスの解析例．下側の矢印はシステムの予測を表す．破線はシステムが予測を誤った例であり，上側にシステムが正しく予測できなかった正解例を示す．

(16) あたしは３週間の退院から帰ってきて調子は良くないです。ところで入院中に１５歳の女の子とお友達になりました。勉強の事ですごく悩んでました。

図の 3 文のうち 1 文目と 2 文目に注目すると，システムが用言述語項構造解析・体言述語項構造解析・橋渡し照応解析・共参照解析をおおよそ正しく行えていることが分かる．しかし，3 文目については「勉強」と「悩んでいました」のガ格が「女の子」であるにも関わらず,「あたし」と出力している．これは 1 文目

と 2 文目では主題を表す助詞「は」が付いている「あたし」が文の中心だったが，2 文目では主題が「女の子」に入れ替わったためと考えられる．現在のモデルはこういった談話構造を捉えることができない．

CAModel が解析したニュースコーパスの例を図 13 に載せる．システムは「中国が援助」や「基地を建設」など，比較的表層的特徴から予測が可能なものについては正しく解析できている．一方で,「中国の海軍」や「ミャンマーに建設」などは予測できていない．これらは文書の内容に関するより深い理解が必要となるような例であるが，他の関係を考慮すれば正しく解析できる可能性がある．例えば「中国の海軍」は，正しく予測できている「海軍の基地を中国が建設」という関係から類推が可能である．他の誤り例に目を向けると，システムは「海軍が進出」という予測と同時に「海軍に進出」という予測をしている．意味を考慮すれば「進出」のガ格とヲ格が異なるものであることは明らかである．以上から，CAModel は未だ十分に結束性を捉えられているとは言えず，他の関係の情報をより解析に活用していくことが今後の課題と考えられる．

第6章　おわりに

本研究では，日本語テキストにおける結束性を統合的に解析することに焦点を当て，BERT を用いて事態性名詞を含む述語項構造解析・橋渡し照応解析・共参照解析の同時解析を提案した．結果，それぞれのタスクで既存手法のスコアを大幅に更新した．特に，述語項構造の中でも難しいゼロ照応解析においては既存手法を 10 ポイントから 20 ポイント上回る結果が得られた．述語項構造解析は基礎解析の中でも特に難しいタスクであり，形態素解析や構文解析とは異なりこれまで実応用は難しかった．しかし，本研究により実応用に耐えうる程度の結果を得ることができた．実際に本研究の成果は，情報集約システムである因果関係グラフ [42] において利用されている．

本研究ではさらに，各タスクの同時解析における効果を統計的に検証した．結果，共参照解析が他のタスクとは性質が異なり，特別に扱うべきであることを示した．この事実に基づく CorefCAModel を作成したが，共参照解析以外には効果が見られないという結果になった．共参照の情報をより効果的に他のタスクに利用できるようなモデルの考案が今後の課題である．

提案手法により，述語項構造解析全体のスコアは向上したが，述語項構造解析の部分問題である格解析においては唯一既存手法を下回っていた．そこで，既存手法が多くの素性を使用していたことに注目し，係り受けの素性を加えた DepCAModel を提案した．結果，格解析において効果が確認され，本手法にはさらに拡張の余地があることが示された．

最後に，本研究によって用言に対する日本語述語項構造解析の精度は実用可能な段階にまで向上した．しかし，未だ体言に対する述語項構造解析や橋渡し照応解析には大きな改善の余地が存在する．これらタスクについても十分な精度での解析を可能にし，ゆくゆくはテキスト中のあらゆる結束性を捉え，高度に構造化されたテキストを得ることが最終課題である．

謝辞

本研究を進めるにあたり，黒橋禎夫教授には日頃より熱心にご指導していただきました．特に，論文投稿の際には執筆に不慣れな私に辛抱強く付き合ってくださいました．心より御礼申し上げます．

転勤後も研究方針や論文作成に関する助言のみならず，コーパスや関連ツールの整備をしてくださいました早稲田大学河原大輔教授に深く感謝申し上げます．

本研究において多様な視点から助言をくださっただけでなく，効率よく研究が進められるよう研究室の計算機環境を日々維持・改善してくださいました村脇有吾講師に深く感謝申し上げます．

日頃から研究に関する相談や議論に付き合ってくださり，また資料作成など基本的な技術を丁寧に教えてくださいました清丸寛一氏に深く感謝致します．

また，同期の皆様をはじめとして黒橋研究室の方々には様々な面で協力をいただいただけでなく，その研究に対する熱心な姿勢から多くの刺激を得ることができました．厚く御礼申し上げます．

最後に，修士課程を通して多方面から支援をいただきました家族と友人のみなさまに深く感謝します．

参考文献

[1] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *Proceedings of NAACL*, pp. 4171–4186 (2019).

[2] Caruana, R.: Multitask learning, *Machine learning*, Vol. 28, No. 1, pp. 41–75 (1997).

[3] Zhang, Y. and Yang, Q.: A Survey on Multi-Task Learning (2018).

[4] Liu, X., He, P., Chen, W. and Gao, J.: Multi-Task Deep Neural Networks for Natural Language Understanding, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, Association for Computational Linguistics, pp. 4487–4496 (2019).

[5] 柴田知秀, 河原大輔, 黒橋禎夫: BERT による日本語構文解析の精度向上, 言語処理学会 第 25 回年次大会, 名古屋, pp. 205–208 (2019.3.13).

[6] 飯田龍, 柴田知秀, 井之上直也: 日本語書き言葉を対象にした省略・共参照解析の誤り分析, 言語処理学会第 21 回年次大会ワークショップ「自然言語処理におけるエラー分析」, No. 3 (2015).

[7] Omori, H. and Komachi, M.: Multi-Task Learning for Japanese Predicate Argument Structure Analysis, *Proceedings of NAACL*, pp. 3404–3414 (2019).

[8] Shibata, T. and Kurohashi, S.: Entity-Centric Joint Modeling of Japanese Coreference Resolution and Predicate Argument Structure Analysis, *Proceedings of ACL*, pp. 579–589 (2018).

[9] Kurita, S., Kawahara, D. and Kurohashi, S.: Neural Adversarial Training for Semi-supervised Japanese Predicate-argument Structure Analysis, *Proceedings of ACL*, pp. 474–484 (2018).

[10] Sango, M., Nishikawa, H. and Tokunaga, T.: Domain Adaptation in Japanese Predicate-Argument Structure Analysis considering First and Second Person Exophora, *Journal of Natural Language Processing*, Vol. 26, No. 2, pp. 483–508 (2019).

[11] Matsubayashi, Y. and Inui, K.: Distance-Free Modeling of Multi-Predicate Interactions in End-to-End Japanese Predicate-Argument Structure Analysis, *Proceedings of the 27th International Conference on Computational Linguis-*

tics, pp. 94–106 (2018).

[12] Matsubayashi, Y. and Inui, K.: Revisiting the Design Issues of Local Models for Japanese Predicate-Argument Structure Analysis, *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, Taipei, Taiwan, Asian Federation of Natural Language Processing, pp. 128–133 (2017).

[13] Ouchi, H., Shindo, H. and Matsumoto, Y.: Neural Modeling of Multi-Predicate Interactions for Japanese Predicate Argument Structure Analysis, *Proceedings of ACL*, pp. 1591–1600 (2017).

[14] 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*, Beijing, China, pp. 961–970 (2015).

[15] Sasano, R. and Kurohashi, S.: A Probabilistic Model for Associative Anaphora Resolution, *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, Singapore, Association for Computational Linguistics, pp. 1455–1464 (2009).

[16] 笹野遼平, 黒橋禎夫: 自動獲得した名詞関係辞書に基づく共参照解析の高度化, 自然言語処理, Vol. 15, No. 5, pp. 99–118 (2008).

[17] 井之上直也, 飯田龍, 乾健太郎, 松本裕治: 日本語文章における直接照応および間接照応の統合的解析, 情報処理学会創立 50 周年記念全国大会講演論文集 (2010).

[18] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V.: RoBERTa: A Robustly Optimized BERT Pretraining Approach, *CoRR*, Vol. abs/1907.11692 (2019).

[19] Li, T., Jawale, P. A., Palmer, M. and Srikumar, V.: Structured Tuning for Semantic Role Labeling, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, Association for Computational Linguistics, pp. 8402–8412 (2020).

[20] Jiang, Z., Xu, W., Araki, J. and Neubig, G.: Generalizing Natural Language Analysis through Span-relation Representations, *Proceedings of the 58th An-*

nual Meeting of the Association for Computational Linguistics, Online, Association for Computational Linguistics, pp. 2120–2133 (2020).

[21] Hou, Y.: Bridging Anaphora Resolution as Question Answering, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, Association for Computational Linguistics, pp. 1428–1438 (2020).

[22] Wu, W., Wang, F., Yuan, A., Wu, F. and Li, J.: CorefQA: Coreference Resolution as Query-based Span Prediction, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, Association for Computational Linguistics, pp. 6953–6963 (2020).

[23] Yu, J. and Poesio, M.: Multi-task Learning Based Neural Bridging Reference Resolution, *ArXiv*, Vol. abs/2003.03666 (2020).

[24] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u. and Polosukhin, I.: Attention is All you Need, *NIPS2017*, pp. 5998–6008 (2017).

[25] Sennrich, R., Haddow, B. and Birch, A.: Neural Machine Translation of Rare Words with Subword Units, *ACL2016*, pp. 1715–1725 (2016).

[26] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R. and Le, Q. V.: XLNet: Generalized Autoregressive Pretraining for Language Understanding, *Advances in Neural Information Processing Systems* (Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. and Garnett, R.(eds.)), Vol. 32, Curran Associates, Inc., pp. 5753–5763 (2019).

[27] Joshi, M., Chen, D., Liu, Y., Weld, D. S., Zettlemoyer, L. and Levy, O.: SpanBERT: Improving Pre-training by Representing and Predicting Spans, *Transactions of the Association for Computational Linguistics*, Vol. 8, pp. 64–77 (2020).

[28] Kurohashi, S. and Nagao, M.: Building a Japanese parsed corpus, *Treebanks*, Springer, pp. 249–260 (2003).

[29] Kawahara, D., Kurohashi, S. and Hasida, K.: Construction of a Japanese Relevance-tagged Corpus, *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, European Language Resources Association (ELRA) (2002).

[30] 萩行正嗣, 河原大輔, 黒橋禎夫: 多様な文書の書き始めに対する意味関係タグ

付きコーパスの構築とその分析, 自然言語処理, Vol. 21, No. 2, pp. 213–248 (2014.4).

[31] Hangyo, M., Kawahara, D. and Kurohashi, S.: Building a Diverse Document Leads Corpus Annotated with Semantic Relations, *Proceedings of PACLIC*, pp. 535–544 (2012).

[32] 黒橋・河原研究室: 日本語構文解析システム KNP version 4.1 使用説明書, https://github.com/ku-nlp/knp/blob/master/doc/manual.pdf.

[33] Morita, H., Kawahara, D. and Kurohashi, S.: Morphological Analysis for Unsegmented Languages using Recurrent Neural Network Language Model, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, Association for Computational Linguistics, pp. 2292–2297 (2015).

[34] Tolmachev, A., Kawahara, D. and Kurohashi, S.: Juman++: A Morphological Analysis Toolkit for Scriptio Continua, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Brussels, Belgium, pp. 54–59 (2018).

[35] 黒橋禎夫: 日本語構文解析システム KNP version 2.0 b6 使用説明書 (1998). 京都大学大学院 情報学研究科.

[36] Kurohashi, S. and Nagao, M.: A syntactic analysis method of long Japanese sentences based on the detection of conjunctive structures, *Computational Linguistics*, Vol. 20, No. 4 (1994).

[37] 黒橋禎夫, 長尾眞: 並列構造の検出に基づく長い日本語文の構文解析, 自然言語処理, Vol. 1, No. 1 (1994).

[38] 河原大輔, 黒橋禎夫: KNP で付与される feature 一覧, `https://github.com/ku-nlp/knp/blob/master/doc/knp_feature.pdf`.

[39] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q. and Rush, A. M.: HuggingFace's Transformers: State-of-the-art Natural Language Processing, *ArXiv*, Vol. abs/1910.03771 (2019).

[40] Bengio, S., Vinyals, O., Jaitly, N. and Shazeer, N.: Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks, *CoRR*,

Vol. abs/1506.03099 (2015).

[41] Tenney, I., Das, D. and Pavlick, E.: BERT Rediscovers the Classical NLP Pipeline, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, Association for Computational Linguistics, pp. 4593–4601 (2019).

[42] 清丸寛一, 植田暢大, 児玉貴志, 田中佑, 岸本裕大, 田中リベカ, 河原大輔, 黒橋禎夫: 因果関係グラフ: 構造的言語処理に基づくイベントの原因・結果・解決策の集約, 言語処理学会 第 26 回年次大会, 茨城 (2020.3).

[43] Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, *CoRR*, Vol. abs/1412.6980 (2014).

[44] Chen, C. and Ng, V.: Chinese Zero Pronoun Resolution with Deep Neural Networks, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 778–788 (2016).

[45] Radford, A., Narasimhan, K., Salimans, T. and Sutskever, I.: Improving language understanding with unsupervised learning, Technical report, OpenAI (2018).

[46] Hou, Y.: Enhanced Word Representations for Bridging Anaphora Resolution, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, New Orleans, Louisiana, Association for Computational Linguistics, pp. 1–7 (2018).

[47] Iida, R., Torisawa, K., Oh, J.-H., Kruengkrai, C. and Kloetzer, J.: Intra-Sentential Subject Zero Anaphora Resolution using Multi-Column Convolutional Neural Network, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1244–1254 (2016).

[48] Lyu, C., Cohen, S. B. and Titov, I.: Semantic Role Labeling with Iterative Structure Refinement, *Proceedings of EMNLP-IJCNLP*, pp. 1071–1082 (2019).

[49] Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K. and Zettlemoyer, L.: Deep Contextualized Word Representations, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long*

Papers), pp. 2227–2237 (2018).

[50] Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K. and Zettlemoyer, L.: Deep Contextualized Word Representations, *EMNLP2018*, pp. 2227–2237 (2018).

[51] Sasano, R., Kawahara, D. and Kurohashi, S.: Automatic Construction of Nominal Case Frames and its Application to Indirect Anaphora Resolution, *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, Geneva, Switzerland, COLING, pp. 1201–1207 (2004).

[52] 笹野遼平, 河原大輔, 黒橋禎夫: 名詞格フレーム辞書の自動構築とそれを用いた名詞句の関係解析, 自然言語処理, Vol. 12, No. 3, pp. 129–144 (2005).

[53] 笹野遼平, 河原大輔, 黒橋禎夫: 自動獲得した知識に基づく統合的な照応解析, 言語処理学会 第 12 回年次大会, pp. 480–483 (2006.3).

[54] 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*, Chiang Mai, Thailand, pp. 758–766 (2011).

[55] 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).

[56] Strubell, E., Verga, P., Andor, D., Weiss, D. and McCallum, A.: Linguistically-Informed Self-Attention for Semantic Role Labeling, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, pp. 5027–5038 (2018).

[57] Imamura, K., Higashinaka, R. and Izumi, T.: Predicate-Argument Structure Analysis with Zero-Anaphora Resolution for Dialogue Systems, *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, Dublin, Ireland, Dublin City University and Association for Computational Linguistics, pp. 806–815 (2014).

[58] 飯田龍, 乾健太郎, 松本裕治, 関根聡: 最尤先行詞候補を用いた日本語名詞句同一指示解析, 情報処理学会論文誌, Vol. 46, No. 3, pp. 831–844 (2005).