

外界一人称と二人称を考慮する 日本語述語項構造解析の分野適応

珊瑚 彩主紀^{1,a)} 西川 仁^{1,b)} 徳永 健伸^{1,c)}

概要：本稿では、日本語述語項構造解析に分野適応の技術を導入することを提案する。現代日本語書き言葉均衡コーパス (BCCWJ) において、メディアの違いにより項の種類の出現頻度が異なることがわかった。特に外界照応においてその傾向は顕著である。過去の日本語述語項構造解析の多くは、新聞記事コーパスを対象にしているために、この傾向には注目していなかった。この問題に取り組むため、まず、RNN ベースの述語項構造解析器をベースラインとして使用し、3種類の異なる分野適応技術とその組み合わせを導入した。現代日本語書き言葉均衡コーパス (BCCWJ) を用いた評価実験では、述語項構造解析には分野依存性があることがわかった。特にガ格の解析では、分野適応が日本語述語項構造解析の性能向上に有効であることが確認され、プログテキストの解析においてベースラインと比較し最大で、2.4% 精度が向上した。

キーワード：日本語述語項構造解析, ゼロ照応, 外界照応, 分野適応

1. 導入

日本語述語項構造解析は、対象とする述語の各格の項を予測するタスクである。述語項構造解析は、様々な自然言語処理アプリケーションの土台となる技術である。本研究が対象とする日本語のような談話指向言語では、文から項が省略されることが多い (Kayama, 2003)。これらの省略された項は、ゼロ代名詞またはテキストの外 (外界) にあるとみなされる。

(1) メールを 書いて v_1 送ったよ v_2 . 読んでね v_3 .

例えば、例 (1) は、3つの述語 (v_1 , v_2 , v_3) と1つの明示的な項候補 (メール) を含んだテキストである。例 (1) を述語項構造解析した結果は表 1 のようになる。ここで、角括弧で囲まれた要素は外界照応、丸括弧は文内ゼロ照応、二重丸括弧は文間ゼロ照応である。 v_1 のヲ格の項である「メール」は、格標識「を」によって明示的に示されており、 v_1 との係り受け関係を持っている。このような名詞は、括弧をつけずに示している。

日本語述語項構造解析は、意味役割付与 (Zhou and Xu, 2015; He et al., 2017) タスクと類似しているが、ゼロ代名詞の照応解析と、表 1 において角括弧で示されている外界

照応の同定まで行う点において異なる。日本語述語項構造解析は、単語が省略されうるという点において、中国語やトルコ語、またロマンス語であるスペイン語、ポルトガル語のような null-subject 言語におけるゼロ照応解析と類似している (Iida and Poesio, 2011; Rello et al., 2012; Chen and Ng, 2016; Yin et al., 2017)。

述語 \ 格	ガ格	ヲ格	ニ格
v_1 : 書いて	[書き手]	メール	[読み手]
v_2 : 送った	[書き手]	(メール)	[読み手]
v_3 : 読んで	[読み手]	((メール))	none

表 1 例 (1) の述語項構造解析結果

過去の日本語述語項構造解析の研究では、形態素及び、構文解析から得られた様々な特徴を利用している (Matsubayashi and Inui, 2017; Hayashibe et al., 2011; Imamura et al., 2014; Shibata et al., 2016; Ouchi et al., 2015; Yoshikawa et al., 2013; Taira et al., 2008)。近年のアプローチでは、中間解析を必要としない end-to-end の手法による解析もある (Ouchi et al., 2017)。

本論文の日本語述語項構造解析への貢献は大きく2つある。第一に、外界照応を3つのサブカテゴリ、つまり、書き手である外界一人称 (exo1)、読み手である外界二人称 (exo2)、その他の外界三人称^{*1} (exoX) に分類する。例 (2)

^{*1} 今回使用したコーパスでは「外界一般」とされているが、本論文

¹ 東京工業大学 情報理工学院
^{a)} sango.m.ab@m.titech.ac.jp
^{b)} hitoshi@c.titech.ac.jp
^{c)} take@c.titech.ac.jp

は、サブカテゴリ化の必要性を示している。

(2) サンドイッチ 食べる。

(私は) サンドイッチを食べる。 /

(あなたは) サンドイッチを食べる？

外界照応の書き手 (exo1) と読み手 (exo2) の両方が、動詞「食べる」の項候補であり、どちらを取るかにより文の意味が変わってくる。これら、2つの意味を区別するために外界照応のサブカテゴリ化が必要である。

第二に、日本語述語項構造解析に分野適応の技術を導入する。Surdeanu et al. (2008) と、Hajič et al. (2009) は訓練データとテストデータの分野 (メディア) が異なると、意味役割付与の性能が低下することを報告している。Yang et al. (2015) は、深層学習手法に分野適応を導入することでこの問題に対して取り組んだ。Imamura et al. (2014) を除いて、日本語述語項構造解析の過去の研究のほとんどが、新聞記事という単一の種類のテキストのみを対象としていたため、分野依存性は問題ではなかった。対話文を解析するために Imamura et al. (2014) は新聞記事を使って述語項構造解析器を訓練した。しかし、その他の種類のメディアについてはまだ調査されていない。我々は様々な種類のメディアのテキストを日本語述語項構造解析の対象とするために、現代日本語書き言葉均衡コーパス (BCCWJ)*2 (Maekawa et al., 2014) を使用した。BCCWJ には、紙媒体として、新聞記事 (PN)、書籍 (PB)、雑誌 (PM)、白書 (OW) といったメディアのテキスト、電子媒体として、インターネット上の QA テキスト (OC)、ブログテキスト (OY) といった様々な種類のメディアから集められたテキストを含んでいる。我々は、約 200 万の単語から構成され、共参照と述語項関係が注釈付けされた BCCWJ のコアデータセットを使用した。次章で詳述するが、外界照応の出現分布はメディアによって大きく異なるため、そのテキストのソースメディアを考慮する必要がある。

我々は、リカレントニューラルネットワーク (Recurrent Neural Network:RNN) ベースのベースモデルから始め、次の 5 種類の分野適応手法を導入し評価実験を行った。(1) Fine-tuning (FT) 手法では、まず、訓練データ全体を用いてモデルを学習させる。学習されたパラメータを初期値とし、ターゲット領域のメディアの訓練データを用いて第 2 段階の学習を行う。(2) Feature augmentation (FA) 手法では、全体で共有されるネットワークと分野固有のネットワークを同時に訓練する (Kim et al., 2016)。分野共通の知識は共有のネットワークで、分野固有の知識は分野固有のネットワークで学習されることを期待している。(3) Class probability shift (CPS) 手法では、分野によって異なる、項の種類毎の、項の出現する確率の事前分布に基づい

では、外界一人称、外界二人称と対比させ、外界三人称と呼ぶこととする。

*2 http://pj.ninjal.ac.jp/corpus_center/bccwj/en/

て、ネットワークが出力する確率にバイアスをかける。(4) Voting (VOT) 手法では、上記 3 つの手法による多数決をとり、出力を決定する。(5) Mixture (MIX) 手法では、上記 (1) から (3) の 3 つの手法を単一のネットワークに組み合わせる。各手法の詳細は、4 節にて説明する。

2. 問題設定

2.1 項種別

日本語述語項構造解析に関する先行研究では様々な種類の項 (項種別) を対象としてきた。表 2 は、先行研究が解析の対象としている項を項種別にまとめたものである。表の列は、言語学的観点から項の種類を分類している。項は、テキスト中に現れるか否かにより、文脈照応か、外界照応かに分けられる。文脈照応は、項が述語と同一文内に出現するか否かにより、文内照応か、文間照応かに更に分けられる。一部の文内照応には、述語と項の係り受け関係がある (intra(dep)) が、必ずしも係り受け関係があるわけではなく、同一文内にあるが、係り受け関係を持たない場合を特に文内ゼロ照応 (intra(zero)) と呼ぶ。文間照応 (inter) も、述語と項は係り受け関係を持たないため、どちらも合わせてゼロ照応と呼ぶ。本論文では外界照応を 3 つのサブカテゴリ、書き手を示す外界一人称 (exo1)、読み手を示す外界二人称 (exo2)、及びその他の外界三人称 (exoX) に分けた。ただし、外界一人称 (exo1)、外界二人称 (exo2) はともに単数のみを扱う。以下では、項種別を示すために表 2 に示すラベルを使用する。ラベル none は、述語がその格に対して、項を取らないことを示す。例えば、自動詞はヲ格をとらない。したがって自動詞のヲ格は none となる。

表 2 から、文間照応よりも、文内照応が盛んに研究されていることがわかる。文間照応の解析は、文内照応と比較して、より広い空間を探索することが必要になるため、より困難な問題といえる。外界照応 exo1 と exo2 の項の解析は、文間照応と異なり、探索空間を大幅に増加させない。次節で詳述するとおり、様々なメディアのテキストに対して、文内述語項構造解析をする際、書き手である exo1、読み手である exo2 まで含めて解析することは重要である。外界照応 exoX の処理には様々な方法が存在する。Imamura et al. (2014) は、exoX を単一のカテゴリを想定しており、Hangyo et al. (2013) は、人かその他で区別している。また、Imamura et al. (2009)、Sasano and Kurohashi (2011)、Hayashibe et al. (2011) は、外界照応と、none を区別していない。これらは、表 2 では、△として示している。本研究では、文内照応と外界照応をともに解析対象とする。exoX と inter は、解析対象の述語は、実際に項を取るが、文内には項が現れていないという点において同じであるため、文内述語項構造解析で解析器が exoX と inter を区別することはできない。そのため、今回は unknown というラベルを付け、一纏めにして扱った。これが我々の研究

研究	\	ラベル	文脈照応			外界照応			
			文内照応		文間照応	exo1	exo2	exoX	none
			係り受け関係有り	ゼロ照応	inter				
			intra(dep)	intra(zero)					
本研究			○	○		○	○	△	○
Matsubayashi and Inui (2017)			○	○					
Ouchi et al. (2017)			○	○					
Shibata et al. (2016)			○	○		○	○		○
Imamura et al. (2014)			○	○	○	○	○	○	○
Hangyo et al. (2013)				○	○	○	○	○	
Yoshikawa et al. (2013)			○	○					
Hayashibe et al. (2011)			○	○	○	△	△	△	△
Sasano and Kurohashi (2011)				○	○	△	△	△	△
Imamura et al. (2009)			○	○	○	△	△	△	△

表2 先行研究が対象としている項種別

で、exoX が △ となっている理由である。まとめると、本研究では、intra(dep), intra(zero), exo1, exo2, none, そして exoX と inter をまとめた unknown を扱う。解析の対象とする述語は、BCCWJにおいて、述語と示されているもののうち、動詞と、事態性名詞とする。

メディア	OC	OY	OW	PB	PM	PN
文長	18.2	16.7	39.6	24.2	19.2	22.6

OC: QA テキスト, OY: ブログテキスト, OW: 白書,
PB: 書籍, PM: 雑誌, PN: 新聞

表3 BCCWJにおける各メディアのテキストの平均文長(形態素数)

2.2 項種別毎の分布による分野依存性

日本語述語項構造解析に関する先行研究では、単一「メディア」のテキストを扱っており、その多くは新聞記事を使用していた。本研究の提案手法の評価には BCCWJ を用い、BCCWJ により定義されている OC, OY, OW, PB, PM, PN の6種のメディアを用いた。メディアによりテキストの特性が異なる可能性があるため、述語項構造解析の性能はメディアの特性の影響を受ける可能性がある。本研究の目的の1つは、メディア固有の特性を考慮するために分野適応の技術を導入し、それが日本語述語項構造解析に有効であることを確認することである。

表3は、BCCWJで定義されている各メディアのテキストの平均文長を示したものである。インターネットコンテンツである OC と OY のテキストは、他のテキストに比べ短い傾向がある。文の長さは、文内述語項構造解析において、項の候補数や、述語と項の距離に影響する。

表4は、6つのメディアに対して格毎に項種別毎の分布の割合を示したものである。OW は、他のメディアとは異

なり、文内ゼロ照応よりも文間ゼロ照応のほうが明らかに少ない。ガ格の exo1 と exo2 の外界照応の分布(影付きの行)は、メディア全体を通して顕著に異なっている。OC の exo1 と exo2 は他のメディアと比較して、かなり高い数値を示している。これは、OC には、対話掲示板形式の QA テキストが含まれており、質問者や回答者としての書き手(exo1)と読み手(exo2)がテキスト中で明示的に言及されないためである。OY テキストも exo1 と exo2 が高い数値となっているが、これは、OC とは異なり、ブログテキストであり、ブログ著者としての書き手(exo1)が話題の中心となることが多いためであると考えられる。OW は、他のメディアとは異なり、外界三人称(exoX)の出現頻度が高い。これは、OW は白書という性質上、組織・集団・団体に対して言及した記述が多く、逆に書き手(exo1)や読み手(exo2)を意識した記述は少ないためであると考えられる。

出版物(PB, PM, PN)は、社会的に関心が高く、客観性のある話題が中心となり、テキスト中に情報が欠損していないことが求められるため、外界照応が出現することは少ない。出版物の中では、PM の書き手(exo1)と読み手(exo2)の出現頻度が比較的高い。これは、今回対象とした出版物の中では、雑誌(PM)が最も著者、読者を意識した記述が多いためであると考えられる。

3. 深層リカレントモデル

我々は、以下の3つの層からなるリカレントニューラルネットワーク(RNN)モデルを用いて、日本語述語項構造解析を実現する。

入力層 単語を特徴ベクトルに変換する。

隠れ層 bi-directional RNN 層と全結合層。

格	述語の \ メディア # 項種別	OC	OY	OW	PB	PM	PN	全体
		16,824	15,612	33,529	32,532	30,410	47,609	176,516
ガ格	none	0.06	0.74	0.19	0.39	0.76	1.00	0.58
	intra(dep)	37.33	35.86	35.75	46.20	43.31	43.06	41.11
	intra(zero)	10.50	13.36	18.20	18.03	14.62	19.75	16.81
	inter	18.25	13.21	8.34	18.46	21.68	18.35	16.58
	exo1	12.13	19.16	0.26	0.69	1.60	0.69	3.49
	exo2	8.26	2.59	0.03	0.32	1.33	0.57	1.46
	exoX	13.46	15.05	37.22	15.88	16.61	16.54	19.93
ヲ格	none	62.73	69.59	45.91	61.29	62.74	59.95	59.13
	intra(dep)	21.43	21.27	41.10	28.41	28.70	31.36	30.37
	intra(zero)	4.57	3.55	5.73	4.66	3.95	3.97	4.45
	inter	7.06	3.59	2.24	3.48	3.07	3.14	3.43
	exo1	0.17	0.25	0.00	0.00	0.03	0.01	0.05
	exo2	0.10	0.02	0.00	0.01	0.01	0.00	0.02
	exoX	3.88	1.68	4.96	2.12	1.41	1.44	2.48
ニ格	none	80.12	87.37	87.12	81.43	85.05	85.71	84.69
	intra(dep)	10.78	9.08	9.03	13.43	12.42	11.23	11.19
	intra(zero)	1.97	1.15	1.65	1.90	1.20	1.44	1.55
	inter	3.14	1.21	0.75	2.41	1.07	1.16	1.49
	exo1	1.42	0.36	0.00	0.02	0.04	0.01	0.18
	exo2	0.75	0.17	0.00	0.02	0.03	0.03	0.11
	exoX	1.81	0.67	1.42	0.79	0.17	0.37	0.78

OC: QA テキスト, OY: ブログテキスト, OW: 白書, PB: 書籍, PM: 雑誌, PN: 新聞

表4 BCCWJにおけるメディア毎の、格毎の項種別の出現分布 (%)

出力層 ソフトマックス関数により、2値分類を行う。

我々のモデルは、各単語にバイナリラベルを出力する。その単語がターゲットの述語に対する解析対象の格の項であるか否かを示すため、それぞれの格に対して別々にモデルを用意する必要がある。図1に、モデルの概要を示す。これは、次のように形式的に表せる。

$$\bar{x} = w_a \oplus w_f \oplus b_f \quad (1)$$

$$h^1 = \text{BiLSTM}(\bar{x}) \quad (2)$$

$$h^2 = \text{linear}(h^1) \quad (3)$$

$$p = \text{softmax}(h^2) \quad (4)$$

我々のモデルは、1文ずつ入力文を受け取る。入力文中の単語 $\{w_t\}_0^T$ は、対応する単語の特徴ベクトル $\{\bar{x}_t\}_0^T$ に変換される。単語の特徴ベクトル \bar{x} は、単語埋め込みベクトル w_a 、品詞埋め込みベクトル w_f 、及び、構文的特徴ベクトル b_f を連結したベクトルとして表現される。特徴ベクトル \bar{x} は、双方向型の Long short-term memory recurrent neural network (BiLSTM) に入力される (Schuster and Paliwal, 1997; Graves et al., 2005)。そして、BiLSTM(\cdot) は、各単語に対して、ベクトル h^1 を計算し出力する。linear(\cdot)

関数は、 h^1 を受け取り、 $h^2 = (h_0^2, h_1^2)$ を出力する。 h_0^2 は単語が述語の項となる確率であり、 h_1^2 は単語が述語の項でない確率である。最後に softmax(\cdot) 関数は、 h^2 を受け取り確率 p を出力する。

3.1 入力層

単語埋め込み、品詞埋め込み、および構文的特徴の3つの特徴を定義する。

3.1.1 単語埋め込み

我々は、Suzuki et al. (2016)^{*3}によって日本語 Wikipedia から作成された単語埋め込みを使用する。

3.1.2 品詞埋め込み

各単語には、最大6層の階層的な品詞タグがつけられている。品詞タグの各層に対して、5次元のランダムベクトルを割り当てた。したがって階層的な品詞タグは、6層のベクトルを連結することによって作る30次元のベクトルによって表す。欠落している層は、ゼロベクトルで埋める。

3.1.3 構文的特徴

構文的特徴ベクトルには、以下の4種類の特徴が含まれている。(1) 単語が各文節において主辞か否かを示す二

^{*3} Japanese Wikipedia Entity Vector http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/

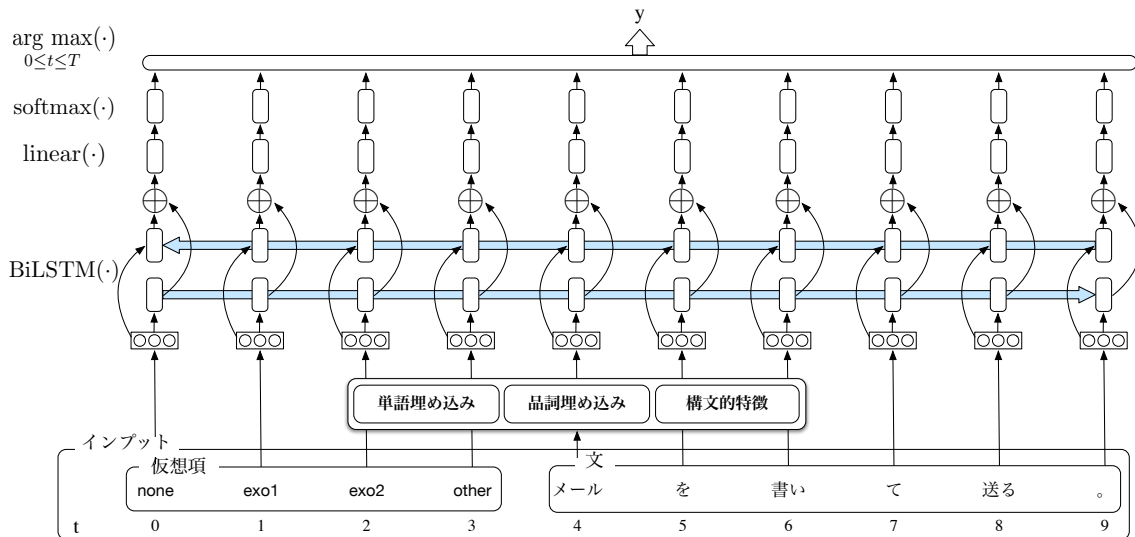


図1 日本語述語項構造解析のための深層リカレントモデル

値ベクトル, (2) コーパスに注釈付けられている文節に基づく, 文頭からの文節距離を示す整数値の特徴ベクトル (入力文の最初の文節の単語は, この値がゼロとなる), (3) 解析されるターゲットの述語からの距離を示す整数値の特徴ベクトル, (4) その単語が解析対象の述語であるか否かを示す二値ベクトルの4種類である.

none, exo1, exo2, そして unknown という4つのラベルを出力するために, 文の先頭の単語の前にそれらを示す仮想項を追加した. これらの仮想項を以下のように割り当てる.

- none** none に対してはゼロベクトルを割り当てた.
- exo1** 「僕」の単語ベクトルを割り当てる. これは, 日本語で一般的な一人称代名詞の一つである.
- exo2** 「お前」の単語ベクトルを割り当てる. これは, 日本語で一般的な二人称代名詞の一つである.
- unknown** 「これ」の単語ベクトルを割り当てる. これは, 日本語で一般的な三人称代名詞の一つである.

日本語には一般的な一人称代名詞, 二人称代名詞が複数存在する. 今回使用した単語埋め込みベクトルは, その性質上, 一人称代名詞と二人称代名詞のコサイン類似度が高く, 「私」と「あなた」のコサイン類似度は, それぞれ語彙集合の中で最も高い. そのため, 「私」と「あなた」を exo1, exo2 の仮想項として採用してしまうと, exo1 と exo2 をうまく区別できない可能性があることを懸念した. 「僕」は一般的な一人称代名詞の中心に近く, また, 二人称代名詞からは比較的遠かった. 「お前」も同様に, 一般的な二人称代名詞の中心に近く, また, 一人称代名詞からは比較的遠かった. そのため, 今回は, 仮想項の exo1, exo2 にそれぞれ「僕」, 「お前」を採用した.

3.2 隠れ層

隠れ層では, 各時刻 t に, 特徴ベクトル \bar{x}_t と h_{t-1}^f を前向き LSTM (LSTM^f) に入力し h_t^f を計算する. 逆に, 各時刻 t に, 特徴ベクトル \bar{x}_t と h_{t+1}^b を後ろ向き LSTM (LSTM^b) に入力し h_t^b を計算する. BiLSTM は, 各時刻 t で h_t^f と h_t^b を連結し, h_t^1 を出力する.

$$h_t^1 = \text{BiLSTM}(\bar{x}_t) = \text{LSTM}^f(\bar{x}_t, h_{t-1}^f) \oplus \text{LSTM}^b(\bar{x}_t, h_{t+1}^b) \quad (5)$$

次に, 2次元ベクトル h_t^2 を得るために, linear(\cdot) 関数に h_t^1 を入力する.

$$h_t^2 = \text{linear}(h_t^1) \quad (6)$$

3.3 出力層

出力層では, 単語が対象とする述語の項であるか否かを判断する. softmax(\cdot) 関数は, 2次元ベクトル h_t^2 を単語がどの程度対象とする述語の項としてふさわしいかを示す確率値に変換する.

$$p_t = \text{softmax}(h_t^2) \quad (7)$$

p_t は時刻 t の単語が項である確率を示す. 我々のモデルは, 最も高い確率 p_y を持つ単語を項として選択する.

$$y = \arg \max_{0 \leq t \leq T} (p_t) \quad (8)$$

4. 分野適応

以下の5つのベースラインモデルを用意した. (1) Each-D モデルは, 単一のメディアのみのデータを使い訓練したモデルである. (2) All モデルは, すべてのメディアのデータを使って訓練したモデルである. (3) Small モデルは, 各メディアの訓練データを75%に減らし, 訓練データのデー

タサイズを小さくしたモデルである。(4) Out-D モデルは、アウトドメインのデータのみを使い訓練したモデルである。(5) One-H モデルは、すべてのメディアを使って訓練したモデルであるが、All モデルとの違いとして、各メディアを示すための 6 次元の one-hot ベクトルが入力層に追加されている。このモデルは、分野適応のベースラインとして用いる。

これらのベースラインモデルに対して、以下の 5 種類の分野適応手法を使ったモデルを用意した。

4.1 (1) Fine-tuning (FT)

1 目目の手法は再学習を行う手法である。すべてのメディアのデータを使ってモデルを訓練し、All モデルを構築する。次に、初期パラメータを All モデルのパラメータにし、ターゲットのメディアのデータのみを使って、訓練することで、各メディアに特化したモデルを構築する。

4.2 (2) Feature augmentation (FA)

2 番目の手法は、すべてのメディアで共通するネットワーク BiLSTM^c に加え、各メディア m に対して固有の BiLSTM^m を用意する手法である Kim et al. (2016)。このモデルの概要は以下ようになる。

$$\bar{x} = w_a \oplus w_f \oplus b_f \quad (9)$$

$$h^1 = \text{BiLSTM}^m(\bar{x}) \oplus \text{BiLSTM}^c(\bar{x}) \quad (10)$$

$$h^2 = \text{linear}^m(h^1) \quad (11)$$

$$p = \text{softmax}(h^2) \quad (12)$$

メディア毎に固有の BiLSTM^m は、そのメディアが固有に持っている特性を学習し、共通の BiLSTM^c は述語項構造解析の一般的な特性を学習することを狙っている。すべてのメディアから無作為に選択したバッチ毎にこのモデルを訓練する。

4.3 (3) Class probability shift (CPS)

3 目目の手法は、各格について、メディア毎の項種別毎の出現分布の違いを反映させる。訓練データ中で、あるメディア m において、解析対象の格として、項種別 tp が出現する確率を p_{tp}^m とする。項種別の確率分布はターゲットメディアによって異なるため、この分布の差を以下のように利用する。各メディア m に対して、2 つの関数 $f^m(h)$ と $g^m(h)$ を次のように定義する。

$$f^m(h) = \frac{p_{tp}^m}{p_{tp}^{\text{All}}} \cdot h \quad (13)$$

$$g^m(h) = \frac{100 - p_{tp}^m}{100 - p_{tp}^{\text{All}}} \cdot h \quad (14)$$

tp は、none, exo1, exo2, unknown, intra のいずれかとなる。ただし、intra ラベルは、文内照応の係り受け有り (intra(dep))

と文内ゼロ照応 (intra(zero)) の両方を含む。

2 次元ベクトル $h^2 = (h_0^2, h_1^2)$ の h_0^2 は単語が述語の項となる確率であり、 h_1^2 は単語が述語の項でない確率である。

$$\bar{x} = w_a \oplus w_f \oplus b_f \quad (15)$$

$$h^1 = \text{BiLSTM}(\bar{x}) \quad (16)$$

$$h^2 = \text{linear}(h^1) \quad (17)$$

$$h^3 = (f^m(h_0^2), g^m(h_1^2)) \quad (18)$$

$$p = \text{softmax}(h^3) \quad (19)$$

式 (18) は、メディア間の項種別の分布を調整することで、出力確率を重みづけする。

4.4 (4) Voting (VOT)

この手法は、上記 3 つの手法の出力の多数決を取る。もし、3 つの手法の出力がすべて異なる場合は、最も確率の高い出力を採用する。

4.5 (5) Mixture (MIX)

最後の手法は、上記 3 つの Fine-tuning, Feature augmentation, Class probability shift を 1 つのモデルとして組み合わせさせたものである。モデルのネットワークは以下の式のように Feature augmentation, Class probability shift を 1 つに組み合わせる。

$$\bar{x} = w_a \oplus w_f \oplus b_f \quad (20)$$

$$h^1 = \text{BiLSTM}^m(\bar{x}) \oplus \text{BiLSTM}^c(\bar{x}) \quad (21)$$

$$h^2 = \text{linear}^m(h^1) \quad (22)$$

$$h^3 = (f^m(h_0^2), g^m(h_1^2)) \quad (23)$$

$$p = \text{softmax}(h^3) \quad (24)$$

このネットワークを使って、Fine-tuning と同じように、すべてのメディアのデータを使ってモデルを訓練し、次に、ターゲットのメディアのデータのみを使って、訓練する。

5. 実験

5.1 実験設定

現代日本語書き言葉均衡コーパス (BCCWJ) を使い評価を行った。各メディア毎にデータを訓練用に 70%、開発用に 10%、テスト用に 20% に分け使用した。各モデルを最大 10 エポック訓練し、開発用データにおいて最も精度が高いモデルを使用した。

5.1.1 ハイパーパラメータ

単語埋め込みと品詞埋め込みの次元数は、それぞれ 200 と 30、BiLSTM のドロップアウト率は 0.2、バッチサイズは 32 とした。我々のモデルは、 α が 0.001、 β が 0.9、weight decay は 0 とし、Adam (Kingma and Ba, 2014) を使い最適化した。Fine-tuning については、weight decay のみ 0.0001 とした。

5.2 実験結果

表5, 表6, 表7は, それぞれガ格, ヲ格, ニ格の各モデルの実験結果を示している. それぞれ, 行見出しがターゲットのメディアを, 列見出しが各モデルを示している.

Each-D モデルは, 訓練データのメディアが精度に与える影響を確認するために用意したモデルである. 訓練データのデータ量がメディアによっては3倍以上違うにもかかわらず, 全体的に, 訓練データとテストデータが同じ(インドメイン)メディアの場合, モデルの精度が高い.

All モデルは, Small モデルよりも精度が高い. これは, 訓練データの量が多い方が精度が向上することを示している. これらの実験結果から, 訓練データの量と, データのメディアの両方を考慮することが必要であるといえる.

All モデルは, すべてのメディアにおいて, Each-D モデルよりも精度が高い. これは, インドメインのデータの他に, アウトドメインの訓練データも学習データとして有効に機能することを示している.

ただし, 訓練データからインドメインのデータを削除すると, 大規模なデータを使用しても性能が低下する. Out-D モデルは, PN^{*4}を除いて, Small モデルよりも訓練データのデータ量が多いにもかかわらず, Small モデルよりも精度が高い. つまり, 訓練データがターゲットメディアのデータを含んでいない場合, データサイズは必ずしも訓練データとテストデータ間でのメディアの不一致を補うとは限らない. 性能を上げるためには, 訓練データにターゲットのメディアのデータが含まれていることが重要であり, インドメイン, アウトドメイン両方のデータをうまく工夫して使う必要がある.

All モデルと One-H モデルを比較すると, 必ずしも One-H モデルの精度が高いというわけではなく, 精度に大きな違いが見られない. つまり, 入力データのメディアを判別するための素性を one-hot ベクトルのような形で与えても, うまくメディアの違いを考慮できないことがわかる.

表5, 表6, 表7の右側は, 分野適応をした結果を示している. 左から順にそれぞれ, Fine-tuning (FT), Feature augmentation (FA), Class probability shift (CPS), Voting (VOT), Mixture (MIX) の結果を示す. 表5のガ格に対する実験結果をみると, VOT モデルは, すべてのメディアにおいて, All モデルよりも精度が高い. VOT モデルと他の分野適応モデルを比較すると, OY のテキストにおいて, CPS モデルは, VOT モデルよりもうまく機能してる. OY のテキストの訓練データのサイズは, 6つのメディアの中で最小であり, このデータの少なさが, OY において VOT の性能が低い原因である可能性がある. 一方, 表6, 表7のヲ格, ニ格に対する実験結果を見ると, 分野適応により一部精度向上が見られるものの, All モデルや One-H モデルの精度

が最も高い場合もある. ベースラインモデルによる実験結果では, ガ格ほど顕著ではないものの, ヲ格, ニ格も, メディア依存性が見られた. だが, 今回使用した分野適応手法はいずれもガ格ほど, うまく働いていない.

表8, 表9, 表10は, ガ格, ヲ格, ニ格それぞれの各メディアにおける項種別毎の精度を示している. 表4によると, OC では exo1 と exo2 が, OY では exo1 が比較的頻出する. そのため, これらの外界照応の精度を個別に分析することは, 正しくそのメディアに適応できたかを考える上で重要である.

表8, 表9, 表10, いずれも, OC の exo1 と exo2, OY の exo1 の箇所(影付きの行)を見ると, 一部, 例外はあるものの All モデルと比べ, 分野適応モデルの精度が向上している. そのため, 分野適応を導入することで, これらの外界照応が出現する際の偏りを解決できたといえる.

表11は, All モデルでは正しく解析されなかったが, VOT モデルにより正しく解析されるようになったガ格の例を示している. ターゲットの述語は太字で示している. OC のテキストには, 対話のような QA のテキストが含まれているので, 最初の例にあるように, 読み手(exo2)は「教えて」のような述語のガ格に当てはまる傾向が高い. OY のテキストには, 書き手が自分の経験や意見を書くブログのテキストが含まれている. そういった場合, 表11の2番目の例にあるようにガ格には, 書き手(exo1)が埋まる傾向がある. OW のテキストには, 白書が含まれている. 白書の性質上, 照応先候補として組織・集団・団体が埋まりやすく, それらは外界三人称として登場していることが多い. そのため, 今回の実験設定では, unknown が埋まる傾向がある. PB のテキストには, 書籍が含まれている. 小説のような物語文では, 照応先候補として物語の登場人物が埋まりやすい. 逆に照応先として書き手(exo1)や読み手(exo2)が出てくることは稀である. 例では, 文中に, 述語のガ格となる, 物語の登場人物が出てきていないため, 今回の実験設定では, unknown が正解となる. PM のテキストの例は, 雑誌のキャッチコピーである. キャッチコピーでは, 一般に書き手(exo1)や読み手(exo2)ではない外界一般がガ格となる. そのため, ガ格は unknown となる. PN のテキストには, 新聞記事が含まれている. 新聞記事に出てくるタイトルの見出しテキストでは, 最初の句がガ格を埋める場合などにおいては, その格標識は省略されることがある. 表11の例は, いずれも, 文単体のみを見れば, All モデルの出力も一見間違いではなさそうである. だが, 上記のようにそれぞれのメディアの特性を考えると, 間違いであることがわかる. 我々の分野適応モデルは, これらの例に示されているように, 項の曖昧性が高い文においてメディア別の傾向をうまくとらえることで, 正しく解析することができている.

*4 PN の Out-D モデルの訓練データは, 123,564 - 33,327 = 90,237 である.

\モデル	ベースライン										分野適応				
	Each-D				Small		All	Out-D	One-H	FT	FA	CPS	VOT	MIX	
ターゲット\サイズ	OC	OY	OW	PB	PM	PN	92,674	123,564	-	-	-	-	-	-	
OC	61.2	53.6	51.6	58.8	58.6	58.3	66.9	68.5	62.4	66.9	67.7	65.6	61.4	69.6	67.2
OY	51.7	54.1	47.5	52.3	52.3	52.5	61.1	60.4	58.4	63.2	63.2	61.5	65.6	64.0	62.8
OW	54.4	57.2	74.3	66.1	64.9	67.4	74.5	76.5	70.6	76.8	73.0	76.1	70.2	77.6	77.2
PB	61.6	57.2	66.1	74.0	71.4	71.1	76.8	78.8	75.1	79.2	73.8	77.6	78.3	79.3	78.0
PM	55.0	52.3	66.2	75.5	72.9	72.8	76.3	76.4	75.1	78.8	74.4	78.7	78.0	80.0	77.4
PN	54.1	54.0	63.2	64.3	65.1	69.8	72.1	74.0	71.1	73.6	70.7	72.5	73.9	74.4	73.0

OC: QA テキスト, OY: ブログテキスト, OW: 白書, PB: 書籍, PM: 雑誌, PN: 新聞

表 5 ガ格の実験結果 (精度)

\モデル	ベースライン										分野適応				
	Each-D				Small		All	Out-D	One-H	FT	FA	CPS	VOT	MIX	
ターゲット\サイズ	OC	OY	OW	PB	PM	PN	92,674	123,564	-	-	-	-	-	-	
OC	83.4	79.0	78.2	81.7	80.1	80.1	83.9	85.3	83.6	85.9	85.3	84.8	84.4	86.1	85.9
OY	81.0	82.4	77.6	80.3	84.0	82.9	84.9	85.8	83.9	86.2	83.8	84.7	85.8	85.8	84.7
OW	65.4	64.9	79.8	73.1	72.4	74.3	81.1	82.1	78.4	81.7	77.9	81.8	76.2	82.9	80.6
PB	84.8	83.7	83.9	86.5	85.8	86.1	88.1	88.6	88.0	89.0	86.6	87.8	88.6	88.3	88.8
PM	80.9	81.2	80.3	82.9	84.2	83.8	85.8	86.4	86.2	85.8	83.7	86.0	85.7	86.3	85.4
PN	77.5	78.1	80.1	79.9	81.6	83.8	85.1	85.8	84.4	85.8	83.8	85.3	85.5	85.9	85.2

OC: QA テキスト, OY: ブログテキスト, OW: 白書, PB: 書籍, PM: 雑誌, PN: 新聞

表 6 ヲ格の実験結果 (精度)

\モデル	Baseline										Adaptation				
	Each-D				Small		All	Out-D	One-H	FT	FA	CPS	VOT	MIX	
ターゲット\サイズ	OC	OY	OW	PB	PM	PN	92,674	123,564	-	-	-	-	-	-	
OC	86.8	84.6	84.0	85.9	85.9	85.6	88.7	88.7	87.3	88.9	87.5	87.9	87.8	89.5	88.7
OY	90.7	91.8	91.5	91.0	92.3	91.8	92.4	92.9	91.7	92.3	92.0	92.2	92.4	92.5	92.2
OW	87.6	87.5	90.5	88.8	88.7	89.2	90.7	91.2	90.4	90.9	89.6	90.8	88.6	91.0	90.1
PB	88.3	88.1	87.6	90.3	89.8	89.2	91.0	91.2	90.6	90.8	90.1	90.4	91.0	91.1	91.2
PM	88.3	88.2	87.6	89.0	90.6	90.0	91.4	91.3	91.3	90.8	90.0	90.4	91.1	91.2	91.3
PN	89.5	90.3	90.6	89.7	91.4	91.7	92.1	92.5	91.8	92.4	91.7	91.9	92.3	92.3	92.4

OC: QA テキスト, OY: ブログテキスト, OW: 白書, PB: 書籍, PM: 雑誌, PN: 新聞

表 7 ニ格の実験結果 (精度)

6. 結論

本稿では、日本語の文内述語項構造解析において、外界照応まで扱うために新たな問題設定を定義し、外界一人称(書き手)、外界二人称(読み手)、外界三人称(その他)として区別して扱うための仮想項の導入、及び、効果的な分野適応手法を提案した。そして、我々は、RNN ベースのモデルと3種類の異なる分野適応技術とその組み合わせを導入し、計5つの分野適応方法を提案した。現代日本語書き言葉均衡コーパス(BCCWJ)を用いて行った、6種のメディアに対しての評価実験では、述語項構造解析において分野依存性があることが示された。特にガ格の解析では、

分野適応が性能を改善するために有効であることを確認した。我々の提案した、仮想項、及び、分野適応手法は、今回提案したRNNベースモデル以外にも様々なニューラルネットワークモデルに対して導入可能である。我々は、今後、文間も含めた効率的な述語項構造解析の分野適応と、分野適応手法の改善を行う予定である。

謝辞 問題設定に関する議論について、松林優一郎博士と笹野遼平博士にご意見をいただきました。感謝いたします。

参考文献

Chen Chen and Vincent Ng. 2016. Chinese Zero Pronoun Resolution with Deep Neural Networks. In *Proceedings of the*

述語の \ ターゲット # 項種別 \ モデル	OC					FT	FA	CPS	VOT	MIX
	Each-D	Small	All	Out-D	One-H					
none	-	-	-	-	-	-	-	-	-	-
intra(dep)	73.5	82.5	85.5	82.9	85.4	83.6	84.0	80.8	86.0	84.1
intra(zero)	32.3	45.8	46.3	39.8	45.8	43.8	44.3	32.8	46.5	45.8
exo1	50.4	40.2	36.3	24.0	41.4	44.3	36.6	0.0	45.5	47.8
exo2	40.5	25.8	40.1	21.8	42.5	42.5	50.8	14.3	46.4	36.1
unknown	66.3	75.9	75.4	70.5	68.5	72.6	66.3	83.2	73.5	69.8

述語の \ ターゲット # 項種別 \ モデル	OY					FT	FA	CPS	VOT	MIX
	Each-D	Small	All	Out-D	One-H					
none	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
intra(dep)	62.0	77.6	78.0	77.1	78.3	79.7	76.7	79.1	79.5	76.2
intra(zero)	33.7	41.4	41.7	42.9	45.4	43.9	47.4	44.4	46.1	43.1
exo1	58.4	27.0	24.9	10.1	57.8	49.4	56.8	63.6	37.5	60.1
exo2	9.1	11.4	27.3	15.9	22.7	6.8	27.3	22.7	13.6	20.5
unknown	53.7	78.3	76.1	81.0	60.2	66.2	56.1	63.8	76.3	60.7

述語の \ ターゲット # 項種別 \ モデル	OW					FT	FA	CPS	VOT	MIX
	Each-D	Small	All	Out-D	One-H					
none	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
intra(dep)	79.2	77.8	83.7	77.3	81.5	73.2	80.3	80.2	82.6	81.7
intra(zero)	33.9	32.2	40.7	22.9	34.1	33.2	32.1	26.6	33.1	34.4
exo1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
exo2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
unknown	86.3	88.1	85.0	83.9	89.8	88.0	89.9	79.5	90.9	90.4

述語の \ ターゲット # 項種別 \ モデル	PB					FT	FA	CPS	VOT	MIX
	Each-D	Small	All	Out-D	One-H					
none	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
intra(dep)	79.7	84.7	86.2	85.5	87.1	77.5	86.0	86.1	86.8	85.6
intra(zero)	45.0	51.1	55.9	51.3	54.4	45.2	52.8	54.5	53.8	52.9
exo1	0.0	14.3	0.0	14.3	0.0	0.0	0.0	0.0	0.0	0.0
exo2	0.0	16.0	56.0	24.0	0.0	0.0	0.0	20.0	0.0	0.0
unknown	83.2	80.8	81.6	74.1	83.4	85.6	80.4	81.3	84.0	82.2

述語の \ ターゲット # 項種別 \ モデル	PM					FT	FA	CPS	VOT	MIX
	Each-D	Small	All	Out-D	One-H					
none	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
intra(dep)	77.6	84.2	86.2	85.1	86.6	75.1	85.1	86.4	85.5	82.7
intra(zero)	47.1	51.3	53.6	51.0	52.9	44.2	53.6	51.6	52.1	44.7
exo1	2.4	2.4	3.2	5.5	0.8	1.6	0.8	0.0	0.0	1.6
exo2	0.0	9.7	9.7	9.7	3.2	12.9	12.9	12.9	6.5	0.0
unknown	81.4	82.0	79.7	78.5	85.0	87.7	85.8	83.9	88.9	87.8

述語の \ ターゲット # 項種別 \ モデル	PN					FT	FA	CPS	VOT	MIX
	Each-D	Small	All	Out-D	One-H					
none	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
intra(dep)	76.3	81.0	84.0	82.5	83.7	78.3	80.6	83.5	83.3	81.5
intra(zero)	45.7	46.1	51.0	47.9	47.9	43.1	45.7	49.7	47.7	48.0
exo1	0.0	6.1	2.4	7.3	1.2	0.0	8.5	0.0	0.0	1.2
exo2	5.4	2.7	24.3	2.7	5.4	0.0	21.6	13.5	5.4	13.5
unknown	79.5	79.8	78.9	74.2	80.0	81.2	81.4	79.8	82.9	80.7

OC: QA テキスト, OY: ブログテキスト, OW: 白書, PB: 書籍, PM: 雑誌, PN: 新聞

表 8 ガ格の実験結果詳細 (精度)

述語の \ ターゲット # 項種別 \ モデル	OC									
	Each-D	Small	All	Out-D	One-H	FT	FA	CPS	VOT	MIX
none	95.8	96.8	96.4	96.7	95.8	94.4	95.4	96.8	96.0	95.2
intra(dep)	73.1	76.7	80.7	79.7	82.6	79.3	79.3	82.8	81.1	82.1
intra(zero)	27.7	27.0	34.0	28.3	23.9	30.8	28.3	39.0	27.7	30.2
exo1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.0
exo2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	25.0	0.0
unknown	58.4	50.3	55.7	41.9	64.1	70.9	61.7	38.6	66.6	65.5

述語の \ ターゲット # 項種別 \ モデル	OY									
	Each-D	Small	All	Out-D	One-H	FT	FA	CPS	VOT	MIX
none	93.5	93.5	92.4	91.7	94.0	93.2	94.4	93.0	93.6	93.4
intra(dep)	70.3	76.9	82.4	76.5	81.3	73.6	74.5	82.3	79.7	77.8
intra(zero)	22.6	23.3	32.3	33.1	24.1	21.1	27.1	32.3	29.3	27.1
exo1	0.0	0.0	0.0	0.0	16.7	0.0	33.3	0.0	0.0	0.0
exo2	-	-	-	-	-	-	-	-	-	-
unknown	39.0	57.9	59.8	57.2	57.9	57.2	48.4	50.9	57.2	48.4

述語の \ ターゲット # 項種別 \ モデル	OW									
	Each-D	Small	All	Out-D	One-H	FT	FA	CPS	VOT	MIX
none	90.8	91.2	90.4	92.5	91.6	93.3	90.9	90.3	92.4	90.2
intra(dep)	86.9	88.9	89.1	83.1	88.5	81.2	88.0	83.6	89.3	88.3
intra(zero)	27.6	26.5	28.9	18.9	24.7	25.0	27.8	17.4	29.3	32.3
exo1	-	-	-	-	-	-	-	-	-	-
exo2	-	-	-	-	-	-	-	-	-	-
unknown	34.8	39.2	51.8	36.8	46.5	31.6	51.8	20.9	49.7	37.3

述語の \ ターゲット # 項種別 \ モデル	PB									
	Each-D	Small	All	Out-D	One-H	FT	FA	CPS	VOT	MIX
none	96.5	96.9	96.9	97.4	96.3	96.8	95.4	96.7	96.9	97.2
intra(dep)	80.6	86.9	87.5	84.5	88.1	80.0	87.3	88.2	86.6	86.4
intra(zero)	26.0	23.0	25.3	21.3	28.3	23.3	32.0	28.3	29.7	26.7
exo1	-	-	-	-	-	-	-	-	-	-
exo2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
unknown	41.8	39.3	44.7	44.0	54.4	46.5	42.8	39.0	39.3	49.1

述語の \ ターゲット # 項種別 \ モデル	PM									
	Each-D	Small	All	Out-D	One-H	FT	FA	CPS	VOT	MIX
none	95.2	94.4	94.5	93.2	94.2	95.8	95.7	93.2	96.0	94.4
intra(dep)	84.9	88.8	90.0	90.3	88.7	81.8	87.7	90.3	88.4	89.0
intra(zero)	27.2	24.1	31.3	33.8	26.0	23.3	35.7	27.4	30.2	27.7
exo1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
exo2	-	-	-	-	-	-	-	-	-	-
unknown	29.6	47.8	43.9	51.1	48.6	34.9	32.4	48.6	36.9	37.2

述語の \ ターゲット # 項種別 \ モデル	PN									
	Each-D	Small	All	Out-D	One-H	FT	FA	CPS	VOT	MIX
none	90.9	91.7	91.7	91.3	93.3	93.1	93.0	91.8	93.0	91.4
intra(dep)	84.4	86.2	87.2	84.7	86.5	82.7	85.6	87.4	87.0	86.4
intra(zero)	23.1	22.3	25.2	21.2	19.1	19.1	23.6	25.2	24.1	21.2
exo1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
exo2	-	-	-	-	-	-	-	-	-	-
unknown	37.7	45.8	50.2	47.4	44.7	24.9	35.1	41.4	37.7	51.6

OC: QA テキスト, OY: ブログテキスト, OW: 白書, PB: 書籍, PM: 雑誌, PN: 新聞

表9 ヲ格の実験結果詳細 (精度)

述語の \ ターゲット # 項種別 \ モデル	OC					FT	FA	CPS	VOT	MIX
	Each-D	Small	All	Out-D	One-H					
none	97.3	98.2	98.8	99.3	98.8	98.2	96.6	99.2	98.2	97.0
intra(dep)	65.9	75.8	71.7	67.9	73.9	72.8	75.0	74.7	78.0	77.5
intra(zero)	16.7	16.7	15.3	11.1	12.5	16.7	15.3	15.3	16.7	19.4
exo1	54.9	51.0	54.9	0.0	58.8	60.8	54.9	0.0	60.8	64.7
exo2	68.2	9.1	36.4	22.7	50.0	45.5	0.0	31.8	63.6	63.6
unknown	15.8	26.3	22.1	14.7	20.5	4.2	37.9	11.1	26.3	29.5

述語の \ ターゲット # 項種別 \ モデル	OY					FT	FA	CPS	VOT	MIX
	Each-D	Small	All	Out-D	One-H					
none	98.1	97.0	98.0	96.1	97.8	98.2	97.2	97.5	97.6	97.5
intra(dep)	60.9	78.2	73.6	74.7	70.9	62.8	74.7	73.6	75.1	69.4
intra(zero)	7.4	22.2	11.1	18.5	11.1	14.8	18.5	18.5	18.5	22.2
exo1	0.0	7.1	0.0	7.1	7.1	0.0	0.0	7.1	0.0	14.3
exo2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
unknown	6.4	7.9	11.1	27.0	0.0	0.0	1.6	3.2	0.0	7.9

述語の \ ターゲット # 項種別 \ モデル	OW					FT	FA	CPS	VOT	MIX
	Each-D	Small	All	Out-D	One-H					
none	98.7	98.1	98.8	98.5	98.4	99.3	98.0	95.3	98.7	96.3
intra(dep)	59.4	64.1	64.8	59.9	64.1	47.4	68.0	67.7	63.6	68.8
intra(zero)	7.8	10.3	8.6	8.6	11.2	5.2	9.5	11.2	8.6	12.9
exo1	-	-	-	-	-	-	-	-	-	-
exo2	-	-	-	-	-	-	-	-	-	-
unknown	2.9	11.2	5.9	2.4	4.7	0.0	2.4	4.1	2.9	21.8

述語の \ ターゲット # 項種別 \ モデル	PB					FT	FA	CPS	VOT	MIX
	Each-D	Small	All	Out-D	One-H					
none	97.2	97.4	98.4	98.1	97.7	96.4	96.7	97.1	97.6	97.3
intra(dep)	76.3	81.4	76.5	75.8	78.3	79.0	82.6	82.1	81.3	80.2
intra(zero)	7.8	11.4	10.6	7.8	9.9	13.5	10.6	13.5	13.5	12.8
exo1	-	-	-	-	-	-	-	-	-	-
exo2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
unknown	14.2	8.5	6.3	3.4	6.3	15.9	2.8	11.4	5.1	20.5

述語の \ ターゲット # 項種別 \ モデル	PM					FT	FA	CPS	VOT	MIX
	Each-D	Small	All	Out-D	One-H					
none	97.2	96.6	97.5	97.0	97.2	98.3	96.5	96.0	97.6	96.8
intra(dep)	74.7	83.4	77.8	79.4	77.0	65.7	78.1	81.5	78.0	79.8
intra(zero)	15.7	16.5	9.6	20.0	10.4	4.4	12.2	13.9	7.8	17.4
exo1	0.0	0.0	25.0	25.0	25.0	0.0	0.0	25.0	0.0	0.0
exo2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
unknown	3.2	8.9	11.3	6.5	0.8	0.0	2.4	31.5	0.8	19.4

述語の \ ターゲット # 項種別 \ モデル	PN					FT	FA	CPS	VOT	MIX
	Each-D	Small	All	Out-D	One-H					
none	97.9	96.8	97.9	96.9	97.9	98.4	97.2	96.9	97.7	97.1
intra(dep)	67.0	77.9	73.3	75.8	73.0	63.3	73.5	78.5	73.5	77.7
intra(zero)	7.5	9.0	7.5	6.0	6.7	3.7	9.0	9.7	6.7	9.0
exo1	-	-	-	-	-	-	-	-	-	-
exo2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
unknown	4.4	8.1	9.6	2.9	1.5	2.2	1.5	8.8	4.4	9.6

OC: QA テキスト, OY: ブログテキスト, OW: 白書, PB: 書籍, PM: 雑誌, PN: 新聞

表 10 ニ格の実験結果詳細 (精度)

メディア	VOT	All	文
OC	exo2	unknown	視力検査の 良い 方法を 教えて 下さい
OY	exo1	unknown	いっぱい 見て 英語 がんばります
OW	unknown	施設	土砂災害 防止 施設の 整備.
PB	unknown	exo1	イライラと 部屋の 中を 歩いた.
PM	unknown	exo1	大好きな RV で 仕事を しよう
PN	トヨタ	unknown	トヨタ 新体制 発表

OC: QA テキスト, OY: プログテキスト, OW: 白書, PB: 書籍, PM: 雑誌, PN: 新聞

表 11 分野適応により正解となった例 (ガ格)

- 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 778–788.
- Alex Graves, Santiago Fernández, and Jürgen Schmidhuber. 2005. Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition. In *Proceedings of the 15th International Conference on Artificial Neural Networks: Formal Models and Their Applications - Volume Part II, ICANN'05*, pages 799–804, Berlin, Heidelberg. Springer-Verlag.
- Jan Hajič, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task, CoNLL '09*, pages 1–18, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Masatsugu Hangyo, Daisuke Kawahara, and Sadao Kurohashi. 2013. Japanese Zero Reference Resolution Considering Exophora and Author/Reader Mentions. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 924–934. Association for Computational Linguistics.
- Yuta Hayashibe, Mamoru Komachi, and Yuji Matsumoto. 2011. Japanese Predicate Argument Structure Analysis Exploiting Argument Position and Type. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 201–209. Asian Federation of Natural Language Processing.
- Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. 2017. Deep Semantic Role Labeling: What Works and What's Next. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 473–483. Association for Computational Linguistics.
- Ryu Iida and Massimo Poesio. 2011. A Cross-Lingual ILP Solution to Zero Anaphora Resolution. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 804–813. Association for Computational Linguistics.
- Kenji Imamura, Ryuichiro Higashinaka, and Tomoko Izumi. 2014. Predicate-Argument Structure Analysis with Zero-Anaphora Resolution for Dialogue Systems. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 806–815, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.
- Kenji Imamura, Kuniko Saito, and Tomoko Izumi. 2009. Discriminative Approach to Predicate-argument Structure Analysis with Zero-anaphora Resolution. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, ACLShort '09*, pages 85–88, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Yuhko Kayama. 2003. L1 aquisition of Japanese zero pronouns: The effect of discourse factors. In *Proceedings of the 2003 Annual Conference of the Canadian Linguistic Association*, pages 109–120.
- Young-Bum Kim, Karl Stratos, and Ruhi Sarikaya. 2016. Frustratingly Easy Neural Domain Adaptation. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 387–396. The COLING 2016 Organizing Committee.
- Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR*, abs/1412.6980.
- Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den. 2014. Balanced Corpus of Contemporary Written Japanese. *Lang. Resour. Eval.*, 48(2):345–371.
- Yuichiro Matsubayashi and Kentaro Inui. 2017. Revisiting the Design Issues of Local Models for Japanese Predicate-Argument Structure Analysis. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 128–133. Asian

- Federation of Natural Language Processing.
- Hiroki Ouchi, Hiroyuki Shindo, Kevin Duh, and Yuji Matsumoto. 2015. Joint Case Argument Identification for Japanese Predicate Argument Structure Analysis. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 961–970. Association for Computational Linguistics.
- Hiroki Ouchi, Hiroyuki Shindo, and Yuji Matsumoto. 2017. Neural Modeling of Multi-Predicate Interactions for Japanese Predicate Argument Structure Analysis. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1591–1600. Association for Computational Linguistics.
- Luz Rello, Ricardo Baeza-Yates, and Ruslan Mitkov. 2012. Elliphant: Improved Automatic Detection of Zero Subjects and Impersonal Constructions in Spanish. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, EACL '12*, pages 706–715, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Ryohei Sasano and Sadao Kurohashi. 2011. A Discriminative Approach to Japanese Zero Anaphora Resolution with Large-scale Lexicalized Case Frames. In *IJCNLP*.
- Mike Schuster and Kuldip K. Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45:2673–2681.
- Tomohide Shibata, Daisuke Kawahara, and Sadao Kurohashi. 2016. Neural Network-Based Model for Japanese Predicate Argument Structure Analysis. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1235–1244. Association for Computational Linguistics.
- Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez, and Joakim Nivre. 2008. The CoNLL-2008 Shared Task on Joint Parsing of Syntactic and Semantic Dependencies. In *Proceedings of the Twelfth Conference on Computational Natural Language Learning, CoNLL '08*, pages 159–177, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Masatoshi Suzuki, Koji Matsuda, Satoshi Sekine, Naoaki Okazaki, and Kentaro Inui. 2016. Neural Joint Learning for Classifying Wikipedia Articles into Fine-grained Named Entity Types. In *Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation: Posters*, pages 535–544.
- Hiroto Taira, Sanae Fujita, and Masaaki Nagata. 2008. A Japanese Predicate Argument Structure Analysis Using Decision Lists. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '08*, pages 523–532, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Haitong Yang, Tao Zhuang, and Chengqing Zong. 2015. Domain Adaptation for Syntactic and Semantic Dependency Parsing Using Deep Belief Networks. *Transactions of the Association for Computational Linguistics*, 3:271–282.
- Qingyu Yin, Yu Zhang, Weinan Zhang, and Ting Liu. 2017. Chinese Zero Pronoun Resolution with Deep Memory Network. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1309–1318. Association for Computational Linguistics.
- Katsumasa Yoshikawa, Masayuki Asahara, and Yuji Matsumoto. 2013. Jointly Extracting Japanese Predicate-Argument Relation with Markov Logic. *Journal of Natural Language Processing*, 20(2):251–271.
- Jie Zhou and Wei Xu. 2015. End-to-end learning of semantic role labeling using recurrent neural networks. In *ACL (1)*, pages 1127–1137. The Association for Computer Linguistics.