

照応・共参照解析を利用した文章の首尾一貫性の評価

飯田 龍^{†1} 徳永 健伸^{†1}

本稿では談話の首尾一貫性の良さを照応・共参照解析の出力に基づいて評価する手法を提案する。「首尾一貫性の高い文章を書く際は、照応・共参照関係を適切に多用する傾向がある」という考えに基づき、入力となる文章についてどの程度照応・共参照関係を同定できたかという解析結果より首尾一貫性の良さを決定する指標を導入した。この指標の良さを定量的に評価するため、日本語の新聞記事を対象とした評価実験を行った。既存研究である Barzilay ら¹⁾が行った首尾一貫性の高い文章と低い文章の対を順序付けるタスクにおいて、既存研究と比較を行い、提案手法、特に名詞句共参照解析の結果を用いた場合に精度良く首尾一貫性の良さを推定できたことについて報告する。

Automatically Evaluating Text Coherence using Anaphora and Coreference Resolution

RYU IIDA ^{†1} and TAKENOBU TOKUNAGA ^{†1}

We propose a metric for automatically evaluating discourse coherence of a text using the outputs of coreference and zero-anaphora resolution models. According to the idea that “one tends to frequently and appropriately utilise anaphoric or coreference relations when writing a coherent text,” we introduce a metric of discourse coherence based on the weighted frequency of automatically identified coreference and zero-anaphoric relations. We empirically evaluated our metric by comparing it to other existing approaches such as Barzilay et al.¹⁾ using Japanese newspaper articles as a target data set. The results indicate that our metric, especially one based on the outputs of noun phrase coreference resolution, better reflects discourse coherence of texts the baseline model.

^{†1} 東京工業大学 大学院情報理工学研究所

Graduate School of Information Science and Engineering, Tokyo Institute of Technology

1. はじめに

談話の首尾一貫性、つまり節や文といった談話単位の間の意味的な整合性を捉えることは談話理解の研究で必須となる課題である。首尾一貫性をどう捉えるかについてはこれまでにさまざまな研究があり、特に修辞構造理論¹⁴⁾に代表される談話単位間の関係を意味的なつながりの観点から明示的に分類するという談話構造解析の研究が盛んに進められている。これらの研究では研究者による言語的な分析に基づき、談話関係の分類が定義されるが、定義された関係の妥当性についての評価は非常に難しい。

この談話単位間の関係から首尾一貫性を考えるという立場の研究に対し、Barzilay らの entity-grid モデル¹⁾に代表される首尾一貫性のモデル化では文章中出现する談話要素（例えば、名詞句など）が文章中でどのように遷移するかを捉えることで首尾一貫性の良さを捉えている（以後、このモデルを首尾一貫性モデルと呼ぶ）。このモデルでは談話要素が隣接する文間でどのような文法役割で出現するかという手がかりをもとに首尾一貫性の良さを推定している。彼女らが提案した entity-grid の枠組みは複数文書要約の後処理となる文の並び換えなどに利用されており、首尾一貫性を考慮する必要がある他の応用処理においても重要な要素技術となると考えられる。

Barzilay らの評価実験では、英語を対象に首尾一貫性の高い文章（人間が書いた新聞記事など）と首尾一貫性の低い文章の対を入力とし、どの程度首尾一貫性の高いものを選択できるかという評価を行っており、地震と事故のドメインの記事に対して 87.2%と 90.4%の精度を得たと報告している。彼女らの手法は明示的に談話要素が出現する英語などの言語には適用しやすいが、日本語のように頻繁に談話要素が省略される場合にはその遷移を捉えることが難しい。横野ら²¹⁾は Barzilay らの entity-grid の枠組みを日本語に直接適用して同様の問題を解いているが、解析精度は英語の場合よりも低くなっており、明示的に文章中出现している談話要素の遷移を捉えるだけでは英語の場合と同程度の解析精度を得るのは難しいのがわかる。省略された格要素を補完するゼロ照応解析を適用し、その結果に対して同様の首尾一貫性モデルを適用することも考えられるが、ゼロ照応解析の解析精度は、我々の最新の研究成果⁸⁾でも F 値で 0.346 と実用レベルに至っているとはいえない。

このような背景から、本研究では entity-grid とは異なるアプローチで首尾一貫性の良さを推定を試みる。首尾一貫性を捉えるための手がかりにはさまざまなものがあるが、ここでは「首尾一貫性の高い文章を書く際は、照応・共参照関係を適切に多用する傾向がある」という考えに基づき、照応・共参照解析器が同定する照応・共参照関係に基づいて首尾一

貫性の良さを推定する手法を提案する．2節で関連する首尾一貫性モデルについて概観し，3節で照応・共参照解析結果に基づく首尾一貫性モデルについて説明する．4節で提案手法で利用する名詞句共参照解析モデルとゼロ照応解析モデルについて紹介する．次に，5節で訓練・評価用データについて説明し，評価用データを4節で導入した名詞句共参照・ゼロ照応解析モデルの評価結果を6節で述べる．さらに首尾一貫性の評価を行い，既存手法と比較した結果を7節で示し，最後に8節でまとめと今後の課題を述べる．

2. 関連研究

首尾一貫性の評価は文章の生成（特に文や節の並び換え）や自動要約の研究において重要な部分問題として多くの研究者に着目されてきた^{1),2),11)-13),16)}．ここでは，Barzilay らの entity-grid モデル¹⁾ に代表される談話要素のつながりと文章の首尾一貫性の関係を扱う研究に限定して概観する．

既存研究の多くは談話要素間の局所的な結束性を扱うセンタリング理論⁵⁾ で定義された概念を導入しており，例えば，Karamanis らの研究¹¹⁾ では，後ろ向き中心 Cb（一つ前の発話に出現する談話要素集合の中で最も顕現性が高いと判断され，かつ現在参照している発話にも出現している談話要素）がどのくらいの頻度で出現するかをもとに文章の首尾一貫性の良さを推定している．

また，Barzilay らの entity-grid の枠組み¹⁾ では，隣接する文間の談話要素の遷移をあらかじめ決めておいた文法役割のクラスで細分化し，首尾一貫性が良さにどのような種類の遷移が影響するかを捉える．例えば，図1の文章を入力として entity-grid を作成したものが図2となる．図2では各文に出現する談話要素がどのような文法役割で出現したかが記録されており，例えば，S→S への遷移が全体のどのくらいの割合で出現するかといった情報がここから抽出される．首尾一貫性の評価の際は，文が正しい順序で並んでいる各文章に対し，文をランダムに並べた首尾一貫性の低い文章を20種類ずつ自動的に作成し，もとの文章との相対的な首尾一貫性の良さを ranking SVM¹⁰⁾ を用いて順序学習する．評価時には，同様の入力セットを作成し，首尾一貫性の高い文章をどの程度選択できるかで評価を行っている．評価には英語で書かれた地震と事故のドメインの記事を用い，それぞれ87.2%と90.4%の精度で首尾一貫性が高いほうの文章を選択できたと報告している．

横野ら²¹⁾ は，この entity-grid を用いた首尾一貫性のモデルを日本語の文章へ適用した．この際，主題「は」と主語「が」を区別する，遷移の間に出現する接続表現のクラスで遷移の粒度を細分化するなどの工夫を行っている．

S ₁	さきがけの武村正義代表は五日、 地元 の 滋賀県 で 記者会見 し、 今夏 の 参院選 について、党公認候補 として 選挙区 で 八、九人を 擁立 する 方針 を 初めて 表明 した。
S ₂	武村代表 は「改選数三以上の東京、 大阪、神奈川、埼玉、愛知 と、改選数一 の 三、四県 で 公認候補 を 擁立 し、比例区 にも 候補者 を 立てた い」と語った。
S ₃	さきがけは同日、参院選滋賀選挙区 に、県議の奥村展三氏を公認候補 として 擁立 すると 発表 した。

下線部が談話要素を表す．

図1 entity-grid モデルへの入力例

	さきがけ	武村代表	五日	方針	...
S ₁	X	S	X	O	...
S ₂	-	S	-	-	...
S ₃	S	-	X	-	...

S, O, X がそれぞれ主格(は/が)、目的語(を/に)、その他の格で出現した談話要素を表す．“-”でマークされている箇所はその文に対象の談話要素が出現していないことを表す．

図2 図1を入力として作成した entity-grid

また，Lin ら¹³⁾ は文法役割ではなく Penn Discourse Treebank (PDTB)¹⁷⁾ に基づく談話関係を推定し，推定した談話関係に関する談話要素の遷移を用いて entity-grid を作成し，Barzilay らと同様の評価を行っている．この結果，談話関係を利用したほうが文法役割の情報を利用した場合よりも精度が良くなったと報告している．

3. 提案手法: 照応・共参照解析結果を利用した首尾一貫性モデル

2節で述べたように，既存の首尾一貫性を捉える研究では文章に明示的に出現している談話要素が各文においてどのように出現しているかを，文法役割や PDTB の談話関係などの観点で表現し，それらがどのように遷移するかを利用している．これに対し，本研究では「首尾一貫性の高い文章を書く際は，照応・共参照関係を適切に多用する傾向がある」という考えを採用し，首尾一貫性の良さを推定する．具体的には，以下の手順で首尾一貫性の良さを推定する．

- (1) 照応・共参照関係がタグ付与されたコーパスをもとに照応・共参照解析モデルを作成する．
 - (2) 首尾一貫性を推定したい文章に(1)で作成した照応・共参照解析モデルを適用し，その結果から首尾一貫性のスコアを計算する．
- (1)で作成した照応・共参照解析モデルは首尾一貫性の高い文章に付与された照応・共参照関係のタグをもとに学習をおこなっているという点に注意されたい．つまり，ここで作成

されたモデルを首尾一貫性の高い文章に関しては比較的妥当な結果を返すことが期待できるのに対し、首尾一貫性が低い（例えば、照応詞とその先行詞の距離が比較的離れているにもかかわらず、照応詞が省略されている、などの）場合は、首尾一貫性が高い文章をもとに作成された解析モデルを適用しても、首尾一貫性が高い文章に適用した場合ほど照応・共参照関係が適切に解析できるとは限らない。そこで(2)の各文章の首尾一貫性のスコアを照応・共参照解析の解析の個数に基づいて求めることを考える。ただし、適用する照応・共参照解析器によっては首尾一貫性の高い文章と低い文章で同じ個数の解析結果を出力する場合もあり、その結果解析の個数だけではどちらの文章が相対的に首尾一貫性が良いかを見積ることができない。ここでは、解析の個数が同じ場合でも、解析時にモデルが出力するスコアに違いが出ることが想定される。つまり、解析個数が同じであっても、首尾一貫性の高い文章のほうが低い場合よりも解析のスコアの総和が大きくなることを期待できる。この考えに基づいて首尾一貫性のスコアを以下のように定義する。

$$\text{coherence}(T) = \frac{1}{N} \sum_j^N \text{score}(j) \quad (1)$$

$$\text{score}(j) = -\log \max_i P(\text{coref}|i, j) \quad (2)$$

ここで、 T は評価対象となる文章であり、 T に出現する照応詞の候補^{*1}を j 、 T における照応詞の出現総数を N とする。 i は照応詞 j の先行詞の候補であり、 $P(\text{coref}|i, j)$ は照応(共参照)解析モデルが出力する解析のスコアである。

また、このモデルが出力するスコアは entity-grid で扱う談話要素の遷移の情報とは独立であるため、提案するモデルのスコアを entity-grid で扱う素性の1つとして利用することも考えられる。2つの手法の統合については7節で述べる。

4. 提案手法で利用する照応・共参照モデル

3節で提案した首尾一貫性のスコアは文章 T に出現する照応詞(候補) j の集合に対して見積るため、一般的には任意の照応・共参照関係に適用できる。ただし、本研究では名詞句の共参照解析とゼロ照応解析のそれぞれの解析結果のそれぞれが首尾一貫性のスコアの算出と最終的な首尾一貫性の推定にどう影響するかを調査するために、2つの解析結果を独立に

利用して首尾一貫性スコアを計算する。以降で、それぞれの解析モデルの詳細をまとめる。

4.1 名詞句共参照解析モデル

名詞句の共参照解析については談話要素間の推移律を考慮したモデル⁴⁾ やクラスタリングに基づく手法³⁾ など、さまざまな手法が提案されているが、本研究で提案する首尾一貫性のスコアでは単純に照応詞(候補)に対して最も先行詞らしい候補を同定した際の解析スコアさえ参照できればスコアが計算できるため、照応詞と先行詞を同時に同定する Ng ら¹⁵⁾ の手法のような単純な手法でもかまわない。ただし、評価実験において共参照解析の性能と首尾一貫性のスコアの有効性の相関を調査したいので、Ng ら¹⁵⁾ の共参照解析モデルに加え、我々が以前提案した探索先行分類型モデル⁶⁾ も導入し、結果を比較する。この探索先行分類型モデルでは、まず最初に先行詞候補間の相対的な先行詞らしさを分類するトーナメントモデル²²⁾ を用いて最も先行詞らしい候補を決定し、その候補と照応詞候補の対が共参照関係となるか否かを分類するという2段階で共参照解析の問題を解く。我々のこれまでの調査では単純な Ng らの手法よりも探索先行分類型モデルを利用したほうが共参照解析の精度が良いことがわかっており*2、これを比較対象とする。さらに、Denis ら⁴⁾ の共参照解析手法では、照応詞を検出する照応性解析モデルと先行詞同定モデルの結果を整数計画法を用いて最適化しており、この結果も Ng らの手法の解析性能を向上させたと報告されている。そこで、照応性判定のモデルとして単純に照応詞候補の局所文脈を参照した場合と、探索先行分類型モデルのように最も先行詞らしい候補を同定したのちにその情報も参照しながら照応性を判定するモデル2種類を用意し、Denis らの最適化の手法によって共参照解析を行った場合についても比較を行う。

前述の共参照解析の既存研究では学習に SVM²⁰⁾ や C4.5¹⁸⁾ などが利用されているが、以降の比較実験では分類器の性能差による影響を除くため、すべての照応・共参照解析器の学習・分類には最大エントロピーモデルを使用する。具体的には、最大エントロピーモデルの実装の一つである MegaM^{*3}を採用した。

4.2 ゼロ照応解析モデル

名詞句共参照解析に加えて、ゼロ照応解析の結果も3節で説明した首尾一貫性モデルに導入する。ただし、既存のゼロ照応解析のモデル(例えば、飯田ら²³⁾ など)は本研究で利用する NAIST テキストコーパス⁷⁾ 内のゼロ照応の解析精度が F 値で 0.346 であり、別の

*1 6節と7節の評価実験では、文節の主辞の品詞が「名詞」である表現を照応詞の候補とした。

*2 探索先行分類型モデルと Ng らの手法の比較の詳細については文献⁶⁾を参照されたい。

*3 <http://www.cs.utah.edu/~hal/megam/>

表 1 ゼロ照応解析結果の精度の変化

解析数	再現率	精度	F 値
7,593	0.345	0.348	0.346
1,632	0.632	0.566	0.597

タスクに解析結果を利用することを考えた場合に十分な解析精度が得られているとはいい難い。そこで、経験的に解析が困難だと考えられる下記の問題をあらかじめ対象外とすることで、解析対象の事例数は減少するが最終的な解析精度が向上するよう調整を行った*1。

- 格交替（使役・受身）をとまなう述語の格要素の省略。
- 引用や名詞句を修飾する節の中に出現するゼロ代名詞。
- 2文以上離れた位置に先行詞を持つゼロ代名詞。
- ヲ格・二格の省略。

これらのゼロ代名詞の多くは談話の埋め込みなどの理由により、ゼロ代名詞と先行詞の関係がセンタリング理論⁵⁾で議論されているような談話要素の顕現性に基づいて同定できない場合であり、これらの問題の解決には談話の深い理解が必要となるので既存研究で利用されている選択選好の情報などを採用したとしても解析は困難なままである。これらを対象外とし、訓練・評価事例から除外することで、より談話の顕現性に特化したモデルが作成できる可能性がある。

本研究では、Iida ら⁸⁾で提案したモデルをゼロ照応解析モデルとして採用する。このモデルはゼロ照応解析の部分問題である先行詞同定、ゼロ代名詞検出、主語同定の3つを独立に解き、その結果を整数計画法で最適化することで最終的な出力を決定している*2。このモデルに加え、ゼロ代名詞と先行詞を同時に同定するモデル（名詞句共参照の Ng ら¹⁵⁾のモデルに相当）や、ゼロ代名詞検出に探索先行分類型モデルを利用する（つまり、最も先行詞らしい候補を同定したのちにその情報も利用してゼロ代名詞を持つかを判定する）場合についても評価を行い、ゼロ照応解析モデルの違いで最終的な首尾一貫性のスコアがどのように変化するかを見る。

予備実験として、探索先行分類型モデルを組み込んだ Iida ら⁸⁾のモデルを用い、上述の解析対象を限定するという実験設定で5節に示す評価用データ 696 記事の評価を行った。この結果、表 1 に示すように、正解として検出すべきゼロ照応関係は 7,593 事例から 1,632

表 2 NAIST テキストコーパス 1.4β 修正版における照応・共参照関係の個数

データ種別	記事数	総文数	総語数	ゼロ照応	名詞句共参照
訓練用	1,753	24,263	651,986	18,526	10,206
評価用	696	9,287	250,901	7,593	4,396

事例に減り、全体の約 2 割しか解析対象としなくなるが、一方精度は F 値で約 6 割となり、少なくとも単純に全ての問題を解く場合よりは首尾一貫性の評価に利用できる見込みがある。そこで、以後の首尾一貫性の評価では上述で示した基準で問題を限定してゼロ照応解析を適用することとする。

5. 評価データ

本実験の評価にはゼロ照応関係と名詞句共参照関係がタグ付与された新聞記事コーパスである NAIST テキストコーパス²³⁾を利用する。このコーパスには訓練と評価の利用に関する明示的な区別がないため、既存研究^{9),19)}で利用されているデータの分け方にしたがって、訓練用と評価用データに分割する。具体的には報道記事の 1 月 1 日分から 11 日分と社説記事の 1 月から 8 月分を訓練事例に、報道記事の 1 月 14 日から 17 日分と社説記事の 10 月から 12 月分を評価用データとした。訓練用データと評価用データに含まれる名詞句共参照関係とゼロ照応関係の個数を表 2 にまとめる。表 2 より、ゼロ照応関係が名詞句共参照関係と比較して多く出現していることがわかる。この数値からも名詞句共参照の関係を解析し entity-grid を作成しても、談話要素の遷移を適切に捉えることができるとは限らないといえる。

6. 実験 1: 照応共参照解析の評価

まず、4 節で導入した名詞句共参照解析・ゼロ照応解析のモデルの解析精度が首尾一貫性の高い文章と低い文章（文をランダムに並び変えた文章）でどのように変化するかを調査した。実験の際は 5 節で示した訓練用のデータでそれぞれの解析モデルを作成し、評価の際は評価用データそのものを首尾一貫性の高い文章集合とし、評価用データ内の各文章に含まれる文をランダムに並び換えたものを首尾一貫性の低い文章として評価を行う。ただし、首尾一貫性の低い場合と高い場合で文の順序が異なるため、ゼロ照応解析の場合は解析対象を同一文内と 1 文前に限定しているため、その範囲に先行詞が出現する場合を正解とした。このため、文の順序が入れ替えられた文章では結束性の高い文章でゼロ照応関係として解析対象に含まれていたものが、解析対象外になる場合がある。

*1 評価実験を通じて評価用データのタグ付与誤りについても若干の修正を行った。

*2 このモデルの詳細については文献⁸⁾を参照されたい。

表 3 名詞句共参照解析の実験結果

	再現率	精度	F 値
original	0.477	0.792	0.595
random 1	0.409	0.751	0.530
random 2	0.405	0.740	0.523
random 3	0.412	0.746	0.531
random 4	0.413	0.746	0.532
random 5	0.406	0.744	0.525

表 4 ゼロ照応解析の実験結果

	再現率	精度	F 値
original	0.632	0.566	0.597
random 1	0.639	0.507	0.566
random 2	0.665	0.499	0.570
random 3	0.663	0.502	0.572
random 4	0.644	0.506	0.567
random 5	0.641	0.497	0.560

表 5 首尾一貫性モデルの評価

	モデル	精度	N/A
	random	0.500	0
(a)	entity-grid (-coref)	0.673	2
	entity-grid (+coref)	0.707	2
(b)	NP: ANT	0.733	1
	NP: ANT + ANA	0.732	1
	NP: ANT + ANA_SCM	0.761	0
(c)	ZERO: ANT	0.523	325
	ZERO: ANT + ANA	0.517	326
	ZERO: ANT + ANA_SCM	0.631	272
	(a) + (b)	0.782	0
	(a) + (c)	0.729	1
	(a) + (b) + (c)	0.794	0

NP, ZERO はそれぞれ名詞句共参照解析, ゼロ照応解析の結果に基づき首尾一貫性のスコアを算出した結果を表す。ANT は Ng ら¹⁵⁾の手法に相当する。ANT+ANA は Denis ら⁴⁾の手法をもとにスコアを算出した結果であり, ANT+ANA_SCM は照応性判定(もしくはゼロ代名詞検出)のスコアを探索先行分類型モデルで得た解析のスコアに置き換えて Denis らの手法で最適化を行った結果を表す。

表 3 に本実験で利用した共参照解析モデルの中で最も精度の良かった探索先行分類型モデルを, 首尾一貫性の高い文章集合と低い文章集合に適用した際の再現率, 精度, F 値を示す。表の original が首尾一貫性が高い文章に共参照解析器を適用した結果であり, random_n が首尾一貫性の低い文章集合 20 種類のうち, n 番目の文章集合に共参照解析器を適用した結果を表している。表 3 から, 首尾一貫性の高い文章で学習した共参照解析器は首尾一貫性の高い文章集合で最も高い F 値を得ており, 期待通り首尾一貫性の高さが解析精度と相関している*1。この解析精度の差が提案する首尾一貫性のスコアに良い影響を与えることが期待できる。

次に, ゼロ照応について同様の評価を行った結果を表 4 に示す。表 4 より, ゼロ照応解析の場合も首尾一貫性の高い文章を解析した結果が首尾一貫性の低い文章を解析した場合を上回っていることがわかる。

7. 実験 2: 首尾一貫性の評価

次に, 首尾一貫性の自動評価について, 既存研究の代表例である Barzilay らの entity-grid モデル¹⁾と提案手法の比較を行った結果をまとめる。評価対象には表 2 に示した評価用データを利用する。ただし, 1 文だけで構成される記事が訓練用データに 213 記事, 評価用データに 156 記事含まれているため, それらを除外して使用する。この評価では, Barzilay ら¹⁾の評価と同様に首尾一貫性の高い文章(オリジナルの文章)のそれぞれに対し, ランダムに文の順序を並び換えた 20 文章を作成する。文をランダムに並び換えた文章を首尾一貫性の低い文章, オリジナルの文章を首尾一貫性の高い文章とみなし, 評価時にはこれらの 2 つの文章を入力とした場合に, どの程度首尾一貫性の高い文章を選択できるかを評価する。

この評価のベースラインモデルには, 首尾一貫性の高い文章と低い文章が入力として与えられた場合にどちらか一方をランダムに選択するモデルを用いる。また, もう一つのベースラインモデルとして Barzilay らの entity-grid モデルを用いるが, このモデルについては共参照解析を用いて談話要素の遷移を捉えるモデル*2(+coref)と文字列一致を用いて近似的に談話要素の遷移を捉えるモデル(coref)の 2 種類を用意した。entity-grid の作成に用いる文法役割のクラスは横野らの研究²¹⁾を参考に, 主題(係助詞「は」が後接する名詞句), 主語(格助詞「が」が後接する名詞句), 目的語(格助詞「を」もしくは「に」が後接する名詞句), その他(それ以外の名詞句)の 4 種類を採用した。ただし, この評価では談話要素間の遷移のみに着目した手法間の比較が目的であるため, Barzilay らが採用した出現頻度に閾値を設け, 遷移の情報を分割して出現頻度の割合を算出するやり方や横野ら²¹⁾が採用した遷移間に出現する接続表現に基づいた遷移の細分化は採用せずにモデルの実装を行った。また, 名詞句共参照解析とゼロ照応解析の結果から算出した首尾一貫性のスコアを entity-grid モデルの素性の一つとして利用した結果についても調査を行った。

結果を表 5 に示す。精度が各手法の正解率を表し, N/A が首尾一貫性の高い文章と低い

*1 ここでは 20 種類のうち 5 種類のみ結果を掲載しているが, 残りの 15 種類の場合についても同様の結果を得ている。

*2 この共参照解析には探索先行分類型モデルを適用した。

文章にモデルを適用した際に出力結果（首尾一貫性のスコアやSVMの分類器が出力する分離平面からの距離）が同じであった個数を表す，

表5の結果から，entity-gridモデルも含めすべてのモデルはランダムに文章を選択するベースラインモデルよりも精度が良いことがわかる．また，entity-gridモデルのみを用いた結果については共参照解析を用いた場合のほうが精度が良く，約7割の問題について正しく首尾一貫性の良さを推定できていることがわかる．

提案する照応・共参照解析の結果を用いた首尾一貫性のスコアを利用した場合については，特に名詞句共参照解析を利用した場合が精度が良く，Denisら⁴⁾のモデルでかつ照応性判定に探索先行分類型モデルの結果を利用した場合（表5のNP: ANT + ANA_SCM）に非常に高い精度を得ている．これは，図3の例に示すように，名詞句共参照解析の結果を利用した場合は首尾一貫性の低い文章の場合には照応詞と先行詞が離れるなどの要因によりもともと解けていた共参照関係を同定できなくなり，結果的に首尾一貫性のスコアが低くなるため，正しく首尾一貫性の良さを推定できたと考えられる．

一方，ゼロ照応解析をもとに首尾一貫性のスコアを求めた場合，図4に示すように，自動検出したゼロ代名詞に対して間違った先行詞を同定してしまい，その結果首尾一貫性の低い側の文章に高いスコアを割り振ってしまう．例えば，この例では文 S'_3 の動詞「見送る」や「選択する」のガ格は「首相」なのだが，誤って「五十嵐広三官房長官」を先行詞として解析してしまっている．これは文が並び換えられたために S'_3 の直前の文に出現している主題を誤って解析してしまっただけであるが，ゼロ照応解析で使用している選択選好や格助詞などの情報を参照した場合，これを誤った先行詞として判断することは困難である．このようにゼロ照応解析の場合は名詞句共参照解析と比較して間違った先行詞を同定してしまう場合が多く，結果的に表5に示された精度はランダムな場合よりは良いもののentity-gridモデルよりも悪い結果となっている．ただし，表5の(a)+(c)の結果を見てわかるように，ゼロ照応解析の結果に基づく首尾一貫性のスコアをentity-gridで用いる素性の一つとすることで，entity-gridモデル単体の場合と比較して精度が向上しており，同様に共参照解析の結果を利用した場合も精度が良くなっていることがわかる．最後に，本研究で利用したentity-gridモデル，名詞句共参照解析の結果に基づくスコア，ゼロ照応解析の結果に基づくスコアの3種類は個々に独立した観点から首尾一貫性の良さを推定していることになるので，この3つを組み合わせた結果である表5の(a)+(b)+(c)はさらに数値が向上しており，比較対象であったentity-gridモデルと比べて精度が約9ポイントも上回っていることがわかる．

首尾一貫性: 高 (オジリナル)		首尾一貫性: 低 (ランダム)	
coherence(T)=8.973		coherence(T')=2.846	
S_1	十四日午前一時三十五分ごろ、東京都世田谷区代沢二の区道で、乗用車 ₁ が電柱に衝突して逃げた、と一一〇番通報があった。	$S'_1 (= S_4)$	運転していたのは俳優の坂上忍容疑者で、酒酔い状態だったため、同署は道路交通法違反の現行犯で逮捕した。
S_2	北沢署 ₂ であて逃げ事件とみて捜査したところ、前部が壊れた乗用車 ₁ を発見。	$S'_2 (= S_6)$	同乗の女優にけがはなかった。
S_3	逃走したためパトカーが追跡、運転していた男を取り押さえた。	$S'_3 (= S_1)$	十四日午前一時三十五分ごろ、東京都世田谷区代沢二の区道で、乗用車 ₁ が電柱に衝突して逃げた、と一一〇番通報があった。
S_4	運転していたのは俳優の坂上忍容疑者 ₃ で、酒酔い状態だったため、同署 ₂ は道路交通法違反の現行犯で逮捕した。	$S'_4 (= S_2)$	北沢署であて逃げ事件とみて捜査したところ、前部が壊れた乗用車 ₁ を発見。
S_5	坂上容疑者 ₃ は二週間のけが ₄ 。	$S'_5 (= S_5)$	坂上容疑者は二週間のけが。
S_6	同乗の女優にけが ₄ はなかった。	$S'_6 (= S_3)$	逃走したためパトカーが追跡、運転していた男を取り押さえた。

同じ下付き数値を持つ太字部分が同一実体を指す共参照関係として同定された結果を表す。

図3 首尾一貫性の高い文章と低い文章における名詞句共参照解析の具体例

8. おわりに

本稿では，名詞句共参照解析とゼロ照応解析の結果に基づく首尾一貫性のスコアの算出方法を提案し，日本語新聞記事コーパスを対象に首尾一貫性の良さを推定する評価を通じて提案手法の有効性を示した．特に名詞句共参照解析の結果を利用した場合に既存手法であるentity-gridモデルの精度より良い結果を得ることができた．また，3種類の手法を組み合わせた場合には，既存手法と比較して約9ポイント精度が向上し，提案する首尾一貫性のスコアが既存手法であるentity-gridモデルを増強する指標として役立つことがわかった．

本研究の今後の発展の方向性としていくつかの課題が考えられる．まず，この論文に示した評価実験では新聞記事コーパスを利用したが，事実を列挙するだけで伝えるべき内容の順序を重要視しなくても良い記事が含まれるため，全ての記事が文レベルでの首尾一貫性を捉えるべき対象になっているとはいえない．この論文で提案した首尾一貫性のスコアは最終的には小論文など著者がある主張を伝えるための文章の推敲支援に利用する予定なので，

首尾一貫性: 高 (オジリアル)		首尾一貫性: 低 (ランダム)	
coherence(T)=3.369		coherence(T')=4.883	
S ₁	「堀さん、どうか頑張ってください。立候補するからには勝利を。」	S' ₁ (= S ₄)	首相は十三日に訪米から帰国したばかりで、周辺には体調不安や疲労を考慮して札幌行きは見送るべきだと意見が強かった。
S ₂	社会党北海道本部のパーティー出席のため十四日、札幌市を訪れた村山富市首相はあいさつで、北海道知事選に社会党などの推薦で出馬する堀達也前北海道副知事を強い口調で支持した。	S' ₂ (= S ₆)	同行した五十嵐広三官房長官 ₂ は「聞いていて涙が出た」と持ち上げたが、伊東氏に厳しい姿勢を示すことで「新党派結成で政権を揺さぶる新民連をけん制した」との見方も。
S ₃	首相に対しては、社会党を離党して自民党の推薦で知事選に出馬する元側近の伊東秀子衆院議員に同情的、との見方が根強くあっただけに首相は態度を明確にしたようだ。	S' ₃ (= S ₅)	だが、堀氏の事実上の決起集会のパーティー出席を見送れば ₂ 「自社政権づくりに功績があった伊東氏への配慮と受け取られるのは確実」とあって、首相自ら札幌行きを選択した ₂ 。
S ₄	首相 ₁ は十三日に訪米から帰国したばかりで、周辺には体調不安や疲労を考慮して札幌行きは見送るべきだと意見が強かった。	S' ₄ (= S ₃)	首相に対しては、社会党を離党して自民党の推薦で知事選に出馬する元側近の伊東秀子衆院議員に同情的、との見方が根強くあっただけに首相 ₁ は態度を明確にしたようだ。
S ₅	だが、堀氏の事実上の決起集会のパーティー出席を見送れば「自社政権づくりに功績があった伊東氏への配慮と受け取られるのは確実」とあって、首相自ら札幌行きを選択した ₁ 。	S' ₅ (= S ₁)	「堀さん、どうか頑張ってください ₁ 。立候補するからには勝利を。」
S ₆	同行した五十嵐広三官房長官は「聞いていて涙が出た」と持ち上げたが、伊東氏に厳しい姿勢を示すことで「新党派結成で政権を揺さぶる新民連をけん制した」との見方も。	S' ₆ (= S ₂)	社会党北海道本部のパーティー出席のため十四日、札幌市を訪れた村山富市首相はあいさつで、北海道知事選に社会党などの推薦で出馬する堀達也前北海道副知事を強い口調で支持した。

下線部が述語であり、対応する下付き数字を持つ太字部分がその述語の格として選択された名詞(句)を表す。

図4 首尾一貫性の高い文章と低い文章におけるゼロ照応解析の具体例

今後は小論文などの文章を収集したのちに、それらの文章でも同様の結果が得られるかを調査する予定である。また、本研究を含め entity-grid モデルから派生している研究では、同一実体を指す表現が文章中で(局所的に)どのように遷移すれば首尾一貫性が高くなるという点に焦点を当てて研究を進めているが、実際は同一実体を指していない場合でも関連する名詞句が文章に偏在しており、それらをどのような観点で捉えモデルに組み込むかが重要な課題になると考えられる。この点についても今後の課題としたい。

謝辞

本研究は科研費若手研究(A)「談話解析技術に基づいた文章推敲支援」(課題番号: 23680014)の支援を受けた。記して謝意を表す。

参考文献

- 1) Barzilay, R. and Lapata, M.: Modeling Local Coherence: An Entity-Based Approach, *Computational Linguistics*, Vol.34, No.1, pp.1-34 (2008).
- 2) Bollegala, D., Okazaki, N. and Ishizuka, M.: A Bottom-Up Approach to Sentence Ordering for Multi-Document Summarization, *In Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pp.385-392 (2006).
- 3) Cai, J. and Strube, M.: End-to-End Coreference Resolution via Hypergraph Partitioning, *In Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pp.143-151 (2010).
- 4) Denis, P. and Baldridge, J.: Joint Determination of Anaphoricity and Coreference Resolution using Integer Programming, *Proc. of HLT/NAACL*, pp.236-243 (2007).
- 5) Grosz, B.J., Joshi, A.K. and Weinstein, S.: Centering: A framework for modeling the local coherence of discourse, *Computational Linguistics*, Vol.21, No.2, pp. 203-226 (1995).
- 6) Iida, R., Inui, K. and Matsumoto, Y.: Anaphora resolution by antecedent identification followed by anaphoricity determination, *ACM Transactions on Asian Language Information Processing (TALIP)*, Vol.4, No.4, pp.417-434 (2005).
- 7) Iida, R., Komachi, M., Inui, K. and Matsumoto, Y.: Annotating a Japanese Text Corpus with Predicate-Argument and Coreference Relations, *Proceeding of the ACL Workshop 'Linguistic Annotation Workshop'*, pp.132-139 (2007).
- 8) Iida, R. and Poesio, M.: 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 (ACL-HLT 2011)*, pp.804-813 (2011).
- 9) Imamura, K., Saito, K. and Izumi, T.: Discriminative Approach to Predicate-

- Argument Structure Analysis with Zero-Anaphora Resolution, *Proceedings of ACL-IJCNLP, Short Papers*, pp.85–88 (2009).
- 10) Joachims, T.: Optimizing Search Engines Using Clickthrough Data, *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pp.133–142 (2002).
 - 11) Karamanis, N., Poesio, M., Mellish, C. and Oberlander, J.: Evaluating centering-based metrics of coherence using a reliably annotated corpus, *In Proceedings of ACL 2004*, pp.391–398 (2004).
 - 12) Lapata, M.: Probabilistic Text Structuring: Experiments with Sentence Ordering, *In Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pp.545–552 (2003).
 - 13) Lin, Z., Ng, H.T. and Kan, M.-Y.: Automatically Evaluating Text Coherence Using Discourse Relations, *Proceeding of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language TEchnologies (ACL-HLT)*, pp.997–1006 (2011).
 - 14) Mann, W.C. and Thompson, S.A.: Rhetorical Structure Theory: Toward a functional theory of text organization, *Text*, Vol.8, No.3, pp.243–281 (1988).
 - 15) Ng, V. and Cardie, C.: Improving Machine Learning Approaches to Coreference Resolution, *In Proceedings of the 40th ACL*, pp.104–111 (2002).
 - 16) Okazaki, N., Matsuo, Y. and Ishizuka, M.: Improving Chronological Sentence Ordering by Precedence Relation, *In Proceedings of Coling 2004*, pp.750–756 (2004).
 - 17) Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A. and Weber, B.: The Penn Discourse Treebank 2.0, *In Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)* (2008).
 - 18) Quinlan, J.R.: *C4.5: Programs for Machine Learning*, The Morgan Kaufmann Series in Machine Learning, Morgan Kaufmann (1993).
 - 19) Taira, H., Fujita, S. and Nagata, M.: Predicate Argument Structure Analysis Using Transformation Based Learning, *Proceedings of the ACL 2010 Conference Short Papers*, pp.162–167 (2010).
 - 20) Vapnik, V.N.: *Statistical Learning Theory*, Adaptive and Learning Systems for Signal Processing Communications, and control, John Wiley & Sons (1998).
 - 21) 横野光, 奥村学: テキスト結束性を考慮した entity grid に基づく局所的一貫性モデル, 自然言語処理, Vol.17, No.1, pp.161–182 (2010).
 - 22) 飯田龍, 乾健太郎, 松本裕治: 文脈の手がかりを考慮した機械学習による日本語ゼロ代名詞の先行詞同定, 情報処理学会論文誌, Vol.45, No.3, pp.906–918 (2004).
 - 23) 飯田龍, 小町守, 井之上直也, 乾健太郎, 松本裕治: 述語項構造と照応関係のアノテーション: NAIST テキストコーパス構築の経験から, 自然言語処理, Vol.17, No.2, pp.25–50 (2010).