2008－NL－187 (7)
2008／9／24

結束性と首尾一貫性から見たゼロ照応解析

飯田龍　乾健太郎　松本裕治
奈良先端科学技術大学院大学 情報科学研究科
〒 630-0192 奈良県生駒市高山町 8916-5
{ryu-i,inui,matsu}@is.naist.jp

ゼロ照応解析の問題を結束性と首尾一貫性の観点から検討する．結束性の観点からは，Walker[21] のキャッシュモデルの実現方法を検討し，統計的機械学習に基づく実装を提案する．このキャッシュモデルを用いて文間ゼロ照応の先行詞候補削減を試み，評価実験を通じて先行詞同定時に解析対象とする先行詞候補を激減できたことを報告する．また，首尾一貫性の観点からは，含意関係認識で利用される推論知識獲得の手法を照応解析の手がかりとすることで解析精度にどのように影響するかについて調査する．新聞記事を対象に先行詞同定の実験を行い，導入した推論規則が解析に有効に働くことを示す．

Zero-anaphora Resolution from the Perspectives of Cohesion and Coherence

Ryu Iida　Kentaro Inui　Yuji Matsumoto
Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama, Ikoma Nara 630-0192 Japan
{ryu-i,inui,matsu}@is.naist.jp

This paper approaches zero-anaphora resolution in the perspectives of cohesion and coherence. From the perspective of cohesion, we examine how to use the *cache model* addressed by Walker [21], and propose a machine learning-based approach for implementing the cache model. Empirical evaluation is conducted in order to reduce the number of antecedent candidates by the proposed cache model, and this results show that the number of the antecedent candidates of each zero-pronoun is dramatically reduced on the task of antecedent identification. From the perspective of coherence, on the other hand, we investigate whether or not the clues introduced in the area of the automatic inference rules acquisition on entailment recognition improve the performance of anaphora resolution. Through the experiments of the antecedent identification task, we demonstrate the impact of incorporating the inference rules into zero-anaphora resolution.

1 はじめに

結束性や首尾一貫性といった談話を特徴付ける観点から談話の諸現象について考察することは言語理解のための興味深い研究対象であり，これまでにさまざまな研究が報告されている．特に談話要素間の照応関係を扱う解析技術は機械翻訳や情報抽出のような応用分野で必要となる要素技術であるだけでなく，談話理解の達成の度合いを示す試金石であるといえる．ここで**照応関係**とは文章中のある談話要素を他の談話要素が指す関係であり，指す側の談話要素を**照応詞**，指される側の談話要素を**先行詞**という．日本語のような言語では頻繁に照応詞が省略される現象が起こり，この省略された照応詞（**ゼロ代名詞**）の先行詞を同定するゼロ照応解析が情報抽出のような応用処理で必須の処理となる．

本稿では，照応解析，特にゼロ照応解析の問題を結束性と首尾一貫性の観点から検討し，現在照応解析の分野で主に研究が進められている統計的機械学習に基づく解析手法 [16, 12, 6, 22] にどのような貢献できるかを考える．

談話要素間の言語レベルのつながりを特徴付ける結束性については談話の局所的な結束性を捉えるセンタリング理論 [4] をもとにさまざまな研究が進められている．このセンタリング理論で 3 つの制約と 2 つの規則で照応詞と先行詞の解釈の選好を説明するが，基本的には，ある談話単位中で最も顕現性が高い要素が次の談話単位で言語化される場合のつながりの良さを各談話単位の意味的な中心（センター）の局所的な遷移と予測されたセンターと実際に言語化されたセンターの（不）一致から説明す

るものである．言語レベルのつながりの良さの解釈を説明するセンタリング理論は，逆に文章が結束性が高く記述されていることを仮定し，ある談話単位に出現する照応詞に対して局所的に顕現性が最も高くなる先行詞候補を先行詞として決定することにより，照応解析の処理を実現するために利用される．例えば，Walker[20] の日本語のゼロ照応解析の手法では，先行詞候補は「主題 > 主語 > 間接目的語 > 直接目的語 > その他」のような顕現性の序列を手がかりにランキングされ，そのランキングに基づき各文ごとの談話のゼロ代名詞の先行詞を決定する．

センタリング理論では文内に複数の照応詞が存在する場合や前文に先行詞がない場合には適切な解釈を導くことができないという欠点があるため実用的ではない．これに対し，機械学習に基づく手法の多くは複数の照応詞が存在する場合も照応詞ごとに独立に解析を行い，また先行詞が前文より前に存在する場合にも前方文脈全体を探索することで頑健に先行詞を同定することができる．しかし，探索範囲を制限せずに先行詞を探索した場合，文章が非常に長い場合には解析効率が悪く，逆にヒューリスティックに探索範囲を制限した場合には先行詞が解析対象に含まれない可能性がある．

この問題を解決するためには，センタリング理論など談話の結束性を考慮する理論で採用されている談話状況の段階的な保持を考える必要がある．談話状況保持についての議論は Miller[11] の記憶容量の話題やより直接的には Walker[21] のキャッシュモデルが関連する．Walker のキャッシュモデルでは短期記憶に相当するキャッシュと長期記憶であるメインメモリを導入し，参照された談話要素をキャッシュを書き換えることで保持し，キャッシュから溢れた古い情報はメインメモリに退避される．過去の情報を参照する場合にはまずキャッシュから探索し，もしキャッシュ内にみつからない場合にはメインメモリから情報を参照するが，どのようにキャッシュに要素を保持するかについては言及されていない．そこで，本研究ではこのキャッシュモデルの実現例を示し，探索候補の削減について評価を行う．

次に，意味レベルの関連性である首尾一貫性の観点からも照応解析の問題を考える．意味レベルの関連性は修辞構造理論 [10] のような節間の関係に着目したものから，“罪を犯す → 捕えられる → 罰せられる” といった Schank のスクリプト知識 [15] など広範に渡る．本研究ではこのような意味レベルの関連性の中で，近年自動獲得の研究が盛んに行われている含意関係認識のための動詞間の推論関係を陽にゼロ照応解析のための手がかりとして利用することを考える．

本稿の構成は以下の通りである．2 節でキャッシュモデルを利用した先行詞候補削減の手法を提案し，その有効性を検証する．次に 3 節で既存の推論規則の獲得手法を説明，それらの手法をゼロ照応の先行詞同定の問題に適用した結果について報告する．最後に 4 節でまとめる．

2 キャッシュを用いた先行詞候補の削減

照応解析の問題を各照応詞に対して先行詞候補集合から最適な先行詞を同定する問題として扱う場合，照応詞が文章の後方に出現すると，その前方文脈には多くの先行詞候補が出現するため，解析のための探索数が爆発するという問題が起こる．既存研究では，この問題に対して，探索範囲を制限する [23]，品詞や性数一致の情報に基づきあらかじめ候補を選別する [8] といった前処理を行うことにより探索範囲を削減している．しかし，文章の記述スタイルなどにより先行詞を探索すべき範囲は変動し，また照応詞がゼロ代名詞の場合は性数一致の情報は利用できないため，ゼロ照応解析の問題を扱う場合には探索範囲の問題を解決することは簡単ではない．本研究では，この先行詞の探索範囲の問題を，Walker[21] のいうキャッシュを使った先行詞候補削減の問題として考える．

ただし，文献 [7] でも述べたように，ゼロ照応の場合は特に文内に先行詞が出現している場合（文内ゼロ照応）と文を越えて先行詞が出現している場合（文間ゼロ照応）で振舞いが異なり，これらを混合して一つのキャッシュという枠組みで扱うことは困難である．本研究では，特に文間ゼロ照応の場合の先行詞候補のキャッシングを考える．

2.1 候補間のランキング問題に基づくキャッシュモデル

本研究では，キャッシングの枠組みを機械学習を用いたランキングの問題として扱い，キャッシュの更新時にキャッシュに保持すべき上位 N 個を選択する問題として考える．キャッシュを用いた照応解析の処理は概ね以下のようなものを想定している．

1. **キャッシュの初期化**: $C_0 := NULL$，$i := 0$.
2. **対象談話単位の照応解析**: 文 S_i 中のすべての照応詞について文 S_i 中の先行詞候補集合 E_i とキャッシュ C_i 内の要素から先行詞を同定する．
3. **キャッシュの更新**: 要素数 M 個の E_i と要素数 N 個の C_i から要素数 N 個の C_{i+1} を作成する．
4. 後続文脈があれば 1〜3 を繰り返す: $i := i + 1$.

ここで，キャッシュの更新の問題は，図 1 に示すように，現在参照している談話単位（例えば，文 S_i）中に出現している M 個の先行詞候補集合とすでにキャッシュの中に保持されている N 個の要素から

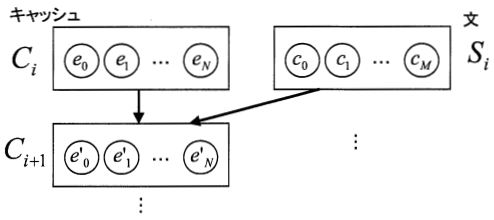

図 1: キャッシュの更新

新たに N 個の要素をキャッシュに保持する問題として考えることができる．この更新時には，後続文脈で出現するゼロ代名詞の先行詞を漏れなく保持しておく必要がある．

この振舞いを学習するために，キャッシュ内の要素もしくは対象文内の先行詞候補が後続文脈中のゼロ代名詞と照応関係になる文ごとに訓練事例を作成する．まず，現在対象としている文中の先行詞候補のうち，後続文脈中のゼロ代名詞と照応関係になる候補を正例，それ以外の候補を負例とする．また，一度でも正例となった候補，つまり一度でもキャッシュの要素として保持された候補は後続文脈のゼロ代名詞のどれからも指されなくなった時点でキャッシュ中に保持する必要がなくなる．この振舞いについても学習する必要があるために，前方文脈で一度でも正例に加えられた各候補は以後訓練事例を作成する文ごとに，後方のゼロ代名詞と照応関係になる候補を正例，それ以外を負例として訓練事例に追加される．

図 2 を例に訓練事例作成について説明する．図中で c_{ij} は文 i の j 番目の候補を表し，ϕ_i が各文に出現するゼロ代名詞を表す．また，矢印が各ゼロ代名詞の先行詞を指す．この状況で訓練事例を作成する場合には，まず文 S_1 で c_{11} と c_{12} が後方のゼロ代名詞 ϕ_i と ϕ_j のそれぞれと照応関係にあるため，正例とする．また，それ以外の候補 c_{13} を負例とする．次に文 S_2 では，後方の ϕ_k と照応関係になる c_{22} を正例に，それ以外を負例とする．前方文脈で一度正例になった c_{11} は後方の ϕ_k と照応関係となるため正例に加え，これに対し，以後照応関係とならない c_{12} は負例に加える．

前述の方法で作成された訓練事例集合を用いて分類器を作成し，実際にキャッシュの更新を行う際には，現在のキャッシュ内の N 個の候補と対象文内に出現する M 個の先行詞候補集合をそれぞれ入力として分類器が出力したスコアの上位 N 個を新規にキャッシュに保持する．このモデルは前文までのキャッシングの状態と現在の文の二つの局所的な文脈に従ってキャッシュを更新するため，以後このモデルを**局所キャッシュモデル**と呼ぶ．

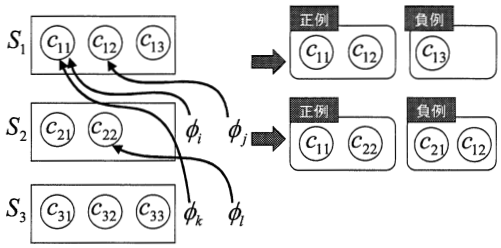

図 2: 局所キャッシュモデルの訓練事例作成の例

局所キャッシュモデルはセンタリング理論 [4] で形式化された局所的な結束性に従ったモデル化の一例であり，このキャッシュの中に保持された候補集合のみを先行詞候補同定の探索範囲とすることで，劇的に探索回数を減少させることができる．ただし，局所キャッシュモデルでは文章の終盤に至るまでに多くのキャッシュの更新を行うため，文章の最初で導入される大域的な主題を表す表現などを誤ってキャッシュから取り除く可能性がある．この問題を回避するために，局所性の情報を捨象したキャッシングについても考える．具体的には，ある文章中が与えられたときにあらかじめ文章中のすべての先行詞候補の先行詞らしさのスコアを求め，各照応詞に関して抽出された先行詞候補集合をそのスコアに基づいてランキングし上位 N 個の候補をキャッシュに保持する．先行詞同定の際にはそのキャッシュ内の候補集合と同一文内のみを探索範囲とすることで，探索数を制限する．以後，このキャッシュモデルを局所キャッシュモデルと対比して，**大域キャッシュモデル**と呼ぶ．

大域キャッシュモデルでは，訓練事例作成時には，文章中のすべてのゼロ代名詞の先行詞となる先行詞候補を正例，それ以外の先行詞候補を負例として訓練事例を作成し，その訓練事例集合をもとに分類器を作成する．先行詞らしさのスコアを求める際は，対象となる文章中のすべての先行詞候補を分類し，各候補に対して分類器が出力した結果をスコアとして利用する．

局所キャッシュモデルが局所的な候補の顕現性を捉えるのに対し，大域キャッシュモデルでは文章の主題を捉えるモデル化を行っていると考えることができるので，お互いのキャッシングの結果は相補的になっている可能性がある．そこで，局所キャッシュモデルに保持された N 個の候補と大域キャッシュモデルに保持された M 個の候補すべてを利用するキャッシュモデルを考える．以後，このモデルを**混合キャッシュモデル**と呼ぶ．

2.2 先行詞候補キャッシングの評価実験

提案する 3 つのキャッシュモデルを用いて，日本語ゼロ照応の問題に対する先行詞候補の削減につ

いて調査を行った．具体的には，評価データ内の各文章に対してキャッシングの処理を行い，文間ゼロ照応のゼロ代名詞が出現した場合に正しく先行詞をキャッシュ内に保持できている割合を調査する．また，異なる文に出現する先行詞候補集合をキャッシングし，その中の候補のみを探索する場合と全探索する場合の探索回数の削減率についても調査する[1]．局所キャッシュモデルと大域キャッシュモデルのキャッシュサイズ（候補保持最大数）は人手で設定する必要があるが，本稿では Miller[11] の記憶領域の議論を参考に 7 個の先行詞候補を保持することにする．また，この局所キャッシュモデルと大域キャッシュモデルのそれぞれの結果を利用することで混合キャッシュモデルの結果を得る．つまり，混合キャッシュモデルでは 14 個の先行詞候補をキャッシュ内に保持することになる．このモデルと直接的に比較を行うため，局所キャッシュモデルと大域キャッシュモデルのキャッシュサイズが 14 の場合も評価する．さらに，評価事例中の 1 文の先行詞候補数の平均が約 7 個であるため，キャッシュサイズの個数 7 個と 14 個に相当する 1 文前までの候補集合と 2 文前までの候補集合を探索した場合をベースラインとする．キャッシュモデルの学習・分類には最大エントロピーモデル[2]を利用した．

訓練事例として NAIST テキストコーパス [26] 中の報道記事 1163 記事，4895 個の文間ゼロ照応となるゼロ代名詞を利用し，評価には報道記事 1157 記事，4365 個の文間ゼロ照応のゼロ代名詞を使用した．また，異なる種類の記述スタイルについてもキャッシングの振舞いを見るために，社説記事 609 記事中の 5231 個の文間ゼロ代名詞についても結果を調査する．評価事例の平均文数は報道記事が 8.07 文，社説記事が 30.72 であり，各ゼロ代名詞について単純に全ての先行詞候補を解析対象とした場合の平均探索数は報道記事が 52.6，社説記事の場合が 100.3 であり，一般に探索すべき候補数は社説記事のほうが多い．

2.2.1 キャッシュングに利用する素性

局所・大域キャッシュモデルの学習・分類で利用する素性は以下の通りである．

- 候補の品詞
- 候補が引用の中に出現しているか否か
- 候補が最初の文に出現したか否か
- 候補助詞の情報 (間接的に主題, 文法役割を表す)
- 候補が格助詞 “は”，“が”，“に”，“を” などを伴った最も直前の候補か否か
- 候補が最後の文節に係る
- ゼロ代名詞から候補までさかのぼったときに出現した接続表現
- キャッシュの中の要素か否か $*$
- 文間の距離 $*$

ただし，$*$ が付いた素性は局所キャッシュモデルでしか利用できない点に注意されたい．

[1] 同一文内の先行詞候補数はキャッシュを用いる場合でも全探索の場合でも変わらないため，削減率の評価対象から除外する．

[2] http://www.cs.utah.edu/~hal/megam/

表 1: キャッシュモデルの評価結果（報道記事）

キャッシュモデル	候補削減率	先行詞のカバー率
baseline(1 文前)	0.149	0.521 (2273/4365)
baseline(2 文前)	0.269	0.713 (3112/4365)
局所モデル (N=7)	0.146	0.850 (3710/4365)
局所モデル (N=14)	0.277	0.915 (3995/4365)
大域モデル (N=7)	0.146	0.748 (3265/4365)
大域モデル (N=14)	0.277	0.851 (3716/4365)
混合モデル (N=M=7)	0.218	0.890 (3886/4365)

2.2.2 キャッシュングの評価結果

実験を行った内容のうち，まず報道記事に関する結果を表 1 に示す．表 1 から，キャッシングを行うことにより，単純に N 文前までの候補を参照した場合と同程度の先行詞候補の削減率を維持しつつ，より多くの先行詞を探索範囲内に保持できていることがわかる．また，局所キャッシュモデルと大域キャッシュモデルの結果を比較することにより，局所的な談話の遷移と文章全体の主題のどちらがより先行詞の保持に適しているかを近似的に見ることができる．結果より，N が 7 の場合も 14 の場合も局所キャッシュモデルのカバー率が大域キャッシュモデルより優れており，先行詞の保持のためには局所的な談話の移り変わりを捉えることが大域的な主題を保持することよりも重要であることがわかる．また，混合キャッシュモデルとそれ以外を比較した場合，局所キャッシュモデルと大域キャッシュモデルの利点を組み合わせた混合キャッシュモデルの結果よりも局所キャッシュモデルの候補保持の性能が優れていることがわかる．これは局所キャッシュモデルのキャッシュサイズが十分に大きいために，局所的な主題はキャッシュ内で必要に応じて入れ替わるが，大域的な主題もキャッシュ内に保持され続けていると考えられる．

次に，社説記事を対象にキャッシングの評価を行った結果を表 2 に示す．前述したように，社説記事のほうが報道記事よりも平均文数が多く，平均探索数が多くなるため，候補を削減する性能は向上するが，先行詞のカバー率は若干品質が悪くなっている．表 2 が示すように，この実験では大域キャッシュモデルはベースラインよりもカバー率が悪く，逆に局所キャッシュモデルはキャッシュサイズが 7 の場合でも 8 割以上の先行詞をキャッシュに保持できていることがわかる．この結果は社説の場合は特に局所的な談話の遷移を捉えなければならないことを示している．

表 2: キャッシュモデルの評価結果（社説記事）

キャッシュモデル	候補削減率	先行詞のカバー率
baseline(1 文前)	0.065	0.566 (2959/5231)
baseline(2 文前)	0.125	0.747 (3910/5231)
局所モデル (N=7)	0.074	0.811 (4240/5231)
局所モデル (N=14)	0.145	0.891 (4662/5231)
大域モデル (N=7)	0.074	0.517 (2702/5231)
大域モデル (N=14)	0.145	0.673 (3523/5231)
混合モデル (N=M=7)	0.126	0.850 (4447/5231)

最後に，具体的なキャッシングの解析例を例 (1) に示す．解析対象となる 4 文目の動詞 “実験する” のガ格ゼロ代名詞に対し，先行詞は 1 文目の “共同研究チーム” である．また，太字箇所が文間の先行詞候補集合であり，括弧内の太字箇所がキャッシュサイズが 7 個の局所キャッシュモデルを用いてキャッシュに保持された先行詞候補である．この例を見てわかるように，単純に 2 文前までの先行詞候補を参照しただけでは解析対象とすることができない先行詞 “共同研究チーム” を適切にキャッシュに保持できていることがわかる．

(1) a. 理論的な**予測**に基づいて**酵素**の**構造**を**一部**変え、**目的**の**化学反応**を起こりやすくする新しい**酵素**を作る**こと**に**世界**で初めて成功したと、**ＮＥＣ**と [**江崎グリコ**] の [**共同研究チーム**$_i$] が**十一日**、発表した。
b. **酵素**は**五千以上**もの**原子**からなり、**構造**が複雑なため、こうした [**理論予測**] は難しかった。
c. 有用な [**化学物質**] を [**効率**] よく作れ、**医薬品**や [**食品など**] の [**分野**] に応用できそうだ。
d. （ϕ_i ガ）実験した のは「ネオプルラナーゼ」という酵素。

3 動詞間の項の共有スコアのゼロ照応解析への利用

3.1 動詞間の項の共有スコア算出のための 3 つの観点

首尾一貫性の情報をゼロ照応解析へ導入する一例として，スクリプト知識のような動詞の連鎖を考える．犯罪のスクリプト知識は，例えば “罪を犯す → 捕えられる → 罰せられる” のような事態の連鎖であり，この構成素である動詞の間では基本的に項が共有される．つまり，この知識を適切に，かつ頑健に構築することができれば，ゼロ解析のためのおおきな手がかりとなる．このような知識を解析に導入しようという試みは，共参照解析の問題に対して Bean ら [2] がすでに行っているが，この試みでは領域を限った評価しか行っておらず，またこの手法を厳密に再現するためには領域ごとに一部の資源を人手で作成する必要がある．一方，一般にスクリプト知識のような資源が頑健にこれらの問題に適用されるためには，少なくとも以下の 2 つの要件を満たす必要がある．

- 領域横断的に利用できる広い知識であること
- 任意の動詞の対に対し，漏れなく推論の適切さのスコアを出力できること

近年注目されている含意関係認識のための推論知識獲得の取り組み [13, 24, 18, 1, 17] は，前述の要件の一つ目，つまり広範な動詞間の知識獲得を目指すものであり，我々の目的実現のために参考とすべきいくつかの特徴を持つ．本稿では，知識獲得のために役立つ 3 種類の文脈情報を考える．

一つ目の文脈情報は，「出現文脈が類似する単語対は類似関係にある」という仮説に基づくもので，例えば，Lin ら [9] の DIRT などがこれに相当する．DIRT では “X is the author of Y ↔ X wrote Y” のような関係を X，Y の文脈の類似性をもとに算出するが，X と Y といった二つの文脈を必要とするため抽出される共起が一般に疎になる．この問題を回避するために，Szpektor ら [17] は文脈を一つに限定した DIRT である unaryDIRT を提案している．彼らは ACE[3] のイベント情報抽出の評価データをもとにした評価手法を用いていくつかの尺度を比較しているが，この結果オリジナルの DIRT に比べ unaryDIRT が精度良くイベント間の関係を捉えていることを示した．そこで，本研究では，この unaryDIRT を動詞間の項の共有スコアの一つとして利用する．unaryDIRT は式 (1) に従う．

$$unaryDIRT(v_i, v_j) = \frac{\sum_{f_{ij}}[MI(f_{ij}, v_i) + MI(f_{ij}, v_j)]}{\sum_{f_i} MI(f_i, v_i) + \sum_{f_j} MI(f_j, v_j)} \quad (1)$$

ここで，f_i は動詞 v_i と共起した文脈を表し，f_{ij} は v_i と v_j の両方と共起した文脈を表す．また，自己相互情報量 MI は式 (2) に従う．

$$MI(v_i, f_i) = log\frac{P(v_i, f_i)}{P(v_i)P(f_i)} \quad (2)$$

次に，2 つ目の文脈情報として鳥澤 [24, 18] が利用した並列構造に着目する．これは，並列関係で出現している 2 つの動詞の間には「典型的にはある時間順序でその動詞が指す事態が連続して起こる」という関係が仮定できるため，並列構造にある動詞の対を収集することで，近似的にスクリプト知識の獲得ができるという考えに基づく．例えば，鳥澤の自動獲得手法 [24] では，任意の名詞に対する推論規則を品質良く獲得する手法を提案している．この手法では，“⟨ 動詞 $_i$ ⟩ て ⟨ 動詞 $_j$ ⟩” や “⟨ 動詞 $_i$(連用形) ⟩ ⟨ 動詞 $_j$ ⟩” のような出現パタンを近似的に並列と見做し，大規模コーパスからこの関係にある動詞対を抽出し，最初に与えられた名詞に関する動詞対の結び付きの強さのスコアを算出している．ただし，このスコアはある名詞に対して計算されるものであ

るため，出力される動詞対は入力となる名詞によって異なる．また，鳥澤はこの知識獲得の結果は得られた推論規則では，動詞対の項が共有されるとは限らないと述べている．並列関係が動詞間の項の共有スコアを求めるためのよい手がかりと考えられる一方，直接的にこの鳥澤 [24, 18] の手法を利用することはできない．そこで，本研究では単純に鳥澤 [24] が利用した “⟨ 動詞 $_i$⟩ て ⟨ 動詞 $_j$⟩” や “⟨ 動詞 $_i$(連用形)⟩⟨ 動詞 $_j$⟩” の 2 種類のパタンを大規模コーパスに適用することで動詞対の頻度を求め，その結果得られた動詞対の共起情報に基づいて自己相互情報量を計算することで 2 つの動詞間の結び付きのスコアとする．ただし，大規模なコーパスから頻度を求めた場合でも頻度が疎になる場合があるので，pLSI[5] を使い，適宜スムージングを行う．

最後に，同一文章内に出現する同一の表現（**アンカー**）が係る動詞の間には何らかの関係があるという仮説に基づき知識獲得を行う取り組みも報告されている．Pekar[13] の手法では，例えば “Mary bought a house. and The house belongs to Mary.” に出現している “Mary” と “house” の 2 つのアンカーを参考に，各アンカーに関連する {buy(obj:X), belong(subj:X)} and {buy(subj:X), belong(to:X)} という 2 つの動詞集合を抽出し，この抽出した動詞集合の構成要素の頻度から動詞対の結び付きの強さを求めている．この手法では，動詞対を抽出する範囲を文の距離もしくは段落の境界で制御するが，その最適な点をあらかじめ知ることは難しい．そこで，本研究では単純に同一文章内からアンカーとなる候補を抽出し[3]，同一アンカーが係る動詞集合中の任意の対を動詞対の頻度として計上する．この頻度に基づき，並列関係の場合と同様に自己相互情報量を計算することで，動詞間の項の共有スコアとする．ただし，後述するように本稿では頻出するガ格のゼロ照応に着目して解析を行うため，アンカーがガ格で動詞に係る場合のみを抽出の対象とした．

3.2 項を共有する述語のランキング問題の評価

3.1 で述べた 3 種のスコア計算のそれぞれ手法がどの程度ゼロ照応解析に影響するかについてゼロ照応の関係がタグ付与されたコーパスを利用して調査する．

まず，述語と述語の関係を見積もることが，先行詞同定にどの程度貢献するかを調査するために，NAIST テキストコーパス [7] 中の 1163 記事から先行詞が述語に係るガ格のゼロ代名詞 5,876 事例を抽出し，動詞間のスコアを利用して先行詞を上位に選出できるかを調査した．対象となるゼロ代名詞の前方に出現する先行詞，もしくはその先行詞を指すゼロ代名詞が係る述語を正解事例，それ以外の候補が係る述語を誤り事例として抽出したところ，5,876 事例中，正解事例は 23,311 事例，誤り事例は事例は 68,612 事例であった．

3.2.1 実験設定

今回利用する 3 種類の動詞対のスコア計算方法では，動詞対を抽出するための大規模コーパスが必要となる．今回の実験ではこのために毎日新聞 1991 年から 1994 年，1996 年から 2003 年までと日経新聞 1990 年から 2000 年までの新聞記事全体を利用し，これらを茶筌[4]と CaboCha[5]を用いて 形態・構文解析して得られた結果から各スコア計算手法に従った動詞対の抽出を行った．また，式 1(1) に従って unaryDIRT を計算するには，どのような文脈 f_i を利用するかを決める必要がある．今回の実験では特にガ格のゼロ照応の問題に着目するため，文脈情報を各動詞のガ格の名詞に限定して自己相互情報量を計算した．この相互情報量の計算にも前述の毎日新聞と日経新聞の 23 年分の全記事を使用し，また頻度が疎になるのを解消するために pLSI[5] を利用して共起の確率を求めた．

実験では，unaryDIRT, 並列関係から抽出した述語の共起スコア（PMI_{coord}），アンカーが係る述語間の共起スコア（PMI_{anchor}）に加え，各動詞対に対してランダムにスコアを出力するベースライン（$random$）とも比較を行う．各スコア付けの手法の評価の結果は式 (3) に基づき評価する．

$$MRR = 1/|N| * \sum_{n \in N} 1/\mathrm{rank(n)} \qquad (3)$$

つまり，各ゼロ代名詞に対して先行詞と関連する動詞をどれだけ上位に出力できたかの平均で評価される．

3.2.2 実験結果

各スコア付け手法の実験結果を表 3 に示す．表 3 より，まず 3 つのスコア付けの手法はベースラインより上位に先行詞と関連する動詞を出力できたことがわかる．また，unaryDIRT より他の 2 手法が優れていることがわかるが，これは unaryDIRT が文章横断的に求めた荒いスコアであるのに対し，並列関係やアンカーを用いたスコア付けは同一文章内から共起を抽出するため，より関連した動詞対を抽出できているのではないかと考えられる．さらに，同一文章内から抽出する 2 つの手法を比較すると，アンカーを用いたほうが結果が良い．これは並列構造

[3] 品詞が “名詞-接尾-*”，“名詞-副詞可能”，“名詞-代名詞-*”，“名詞-非自立-*”，“名詞-数”，“名詞-特殊-*”，“名詞-動詞非自立的”，“名詞-接続詞的”，“名詞-引用文字列” を除くすべての名詞をアンカーの候補とした．

[4] http://chasen.naist.jp/hiki/ChaSen/

[5] http://chasen.org/~taku/software/cabocha/

表 3: 動詞対の共起スコアの有効性の評価

評価尺度	MRR
random	0.543
unaryDIRT	0.565
$MI_{coord}(v_i, v_j)$	0.590
$MI_{anchor}(v_i, v_j)$	**0.621**

で文脈を限定した場合に比べ，アンカーを導入した場合はノイズが少なく今回の問題に適した共起が収集できたためだと考えられるが，この2種の方法では抽出される共起の規模が異なるため，今後抽出対象を変化させた場合の振舞いについても調査する予定である．

3.3 ゼロ照応解析の先行詞同定タスク

次に，ガ格ゼロ代名詞の先行詞同定の問題に動詞対のスコアを導入することで解析精度がどのように影響するかを調査するために，3.1 の各スコアを解析のための素性の一つに加えて比較を行った．

3.3.1 実験設定

先行詞同定の解析モデルには我々が以前提案したトーナメントモデル [25] を使用した．このモデルではゼロ代名詞に対して先行詞候補集合内の各候補間で勝ち抜き戦を行うことで最終的に一つのもっとも先行詞らしい候補を決定する[6]．ただし，実験では2節で導入したキャッシュサイズ 14 の局所キャッシュモデルを利用した．

また，先行詞同定の学習・分類には Support Vector Machine [19]，カーネルには線形カーネルと多項2次カーネルを利用し，パラメタは default 値を用いた[7]．訓練には NAIST テキストコーパス [26] の 1,163 記事中のゼロ代名詞 9,122 事例を使用し，評価には 1,157 記事中の 8,952 事例を使用した．

3.3.2 素性

解析に利用した素性はおおきく以下の4種類に分類できる．

ゼロ代名詞に関する素性：ゼロ代名詞が係る述語が交替を伴って出現している，引用の中に出現している．

先行詞候補に関する素性：先行詞候補の主辞の品詞，格助詞，文頭に出現している，文末に出現している，固有名か否か，候補が格助詞 “は”，“が”，“に”，“を” などを伴った最も直前の候補か否か，これまでに何度先行詞となったか，引用の中に出現している．

ゼロ代名詞と先行詞候補の対に関する素性：大規模コーパスから計算した先行詞候補とゼロ代名詞が係る動詞との自己相互情報量，- ゼロ代名詞が係る述語より候補が先に出現している，ゼロ代名詞が係る述語が候補に直接係る，逆に候補が述語に係る，ゼロ代名詞と候補の距離（どのくらい文数が離れているか）

[6]学習・分類に関する詳細は文献 [25] を参照されたい．

[7]http://www.cs.cornell.edu/people/tj/svm_light/

表 4: 項の共有スコアを用いた先行詞同定の評価

カーネル	スコア無し	スコア有り
線形	0.457 (4091/8952)	0.464 (4157/8952)
多項2次	0.506 (4529/8952)	**0.510** (4562/8952)

先行詞候補対に関する素性：候補間の距離（どのくらい文数が離れているか），先行詞候補とゼロ代名詞が係る動詞との間の自己相互情報量の差

これら4種に加え，動詞対の共起スコアが計算できる場合，つまりゼロ代名詞が動詞に係り，先行詞候補も同様に動詞に係る，もしくは一度他のゼロ代名詞の先行詞となったためそのゼロ代名詞の動詞の情報が保持されている場合は，3.2 の評価で最も結果が良かったアンカーに基づく共起スコアを計算し素性に加える．

3.3.3 実験結果

先行詞同定における動詞対のスコアの導入による精度の変化を表 4 に示す．この表からわかるように，スコアを単純に素性に加えただけでも解析精度が向上しており[8]，動詞対の遷移の情報がゼロ照応解析に役立っていることがわかる．

ここで動詞間の共起スコアを導入した場合にのみ正しく解析できた結果を例 (2) に示す．この例では，“米国$_i$” は動詞 “置かれている” に係り，また動詞 “支持する” のガ格は正しく “米国$_i$” と解析されている．この状況で，動詞 “推進する” のガ格ゼロ代名詞を解析する際には，前方文脈で解析された “米国ガ支持する” と対象となる動詞 “支持する” の動詞対から MI_{anchor}(推進する, 支持する) を計算し，その結果を先行詞同定に利用する．この解析例より，スクリプトとみなすことができる動詞対 “推進する → 支持する” に関して，今回導入した手法が高いスコアを出力し，そのスコアが先行詞同定の処理で有効に働いたために先行詞 “米国$_i$” を適切に選択できたと考えられる．

(2) **米国**$_i$ は米露間の現実的な戦略的利益に立ってエリツィン政権を（ϕ_i ガ）支持せざるを得ず、「エリツィンのジレンマはクリントンのジレンマ」という状況に置かれている ためだ。... ロシアの脅威を骨抜きにした状態で米露核軍縮を（ϕ_i ガ）推進し、同時に旧東欧諸国への北大西洋条約機構拡大を目指している。

4 おわりに

本稿では，結束性と首尾一貫性の観点からゼロ照応の問題を議論した．結束性の観点から，センタリング理論で導入されている考え方を発展し，各談話単位での先行詞候補の棄却問題を考えた．具体的には，候補をキャッシングする問題を機械学習のラン

[8]McNemar 検定，$p < 0.05$ で有意差あり．

キング問題とみなし，キャッシュに保持する候補の最適化を行う手法を提案した．この手法を導入することで，ベースラインと同程度の先行詞候補の削減率を維持しつつ，より多くの先行詞を探索範囲内に含むことができることを示した．また，首尾一貫性の観点からは，含意関係認識の問題で獲得される推論知識をゼロ照応解析の手がかりとすることで，解析精度に影響が出るかを調査した．今回の取り組みは，領域に依存せずに獲得した大規模な知識がどのように照応解析に影響するかを調査した初めての試みであるが，動詞対の共起スコアをガ格ゼロ代名詞の先行詞同定処理に導入することで解析精度が向上するという結果を得た．今回の調査では動詞対の共起スコアを単純に一つの素性として利用しただけだが，今後は文章全体を通じて精度最大化を目指す枠組みの中で利用する予定である．例えば，省略の連鎖を考えた場合，「(φガ)〈動詞$_i$〉,...,(φガ)〈動詞$_j$〉」の間の関係のように省略された要素間で先行詞が同じであるかを陽に解析し，その結果を他の解析結果と組み合せるといった方向性が考えられる．この種の問題は Markov Logic Network[14] などの確率と論理を組み合わせた枠組みの中で形式化できると考えられ，今後の研究課題として取り組む予定である．

参考文献

[1] Abe, S., Inui, K. and Matsumoto, Y.: Two-Phased Event Relation Acquisition: Coupling the Relation-Oriented and Argument-Oriented Approaches, *Proceedings of The 22nd International Conference on Computational Linguistics (COLING)*, pp. 1–8 (2008).

[2] Bean, D. and Riloff, E.: Unsupervised learning of contextual role knowledge for coreference resolution, *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*, pp. 297–304 (2004).

[3] Doddington, G., Mitchell, A., Przybocki, M., Ramshaw, L., Strassel, S. and Weischedel, R.: Automatic Content Extraction (ACE) program - task definitions and performance measures, *Proceedings of the 4rd International Conference on Language Resources and Evaluation (LREC-2004)*, pp. 837–840 (2004).

[4] 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).

[5] Hoffman, T.: Probabilistic latent semantic indexing, *Proceedings of ACM SIGIR*, pp. 50–57 (1999).

[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., Inui, K. and Matsumoto, Y.: Zero-anaphora resolution by learning rich syntactic pattern features, *ACM Transactions on Asian Language Information Processing (TALIP)*, Vol. 6, No. 4 Article. 12 (2007).

[8] Lappin, S. and Leass, H. J.: An Algorithm for Pronominal Anaphora Resolution, *Computational Linguistics*, Vol. 20, No. 4, pp. 535–561 (1994).

[9] Lin, D. and Pantel, P.: Discovery of inference rules for question answering, *Natural Language Engineering*, Vol. 7, No. 4, pp. 343–360 (2001).

[10] 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).

[11] Miller, G. A.: The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information, *The Psychological Review*, Vol. 63, pp. 81–97 (1956).

[12] Ng, V. and Cardie, C.: Improving Machine Learning Approaches to Coreference Resolution, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 104–111 (2002).

[13] Pekar, V.: Acquisition of verb entailment from text, *Proceedings of Human Language Technology Conference/North American chapter of the Association for Computational Linguistics annual meeting (HLT-NAACL06)*, pp. 49–56 (2006).

[14] Richardson, M. and Domingos, P.: Markov logic networks, *Machine Learning*, Vol. 62, No. 1-2, pp. 107–136 (2006).

[15] Schank, R. and Abelson, R.: *Scripts Plans Goals and Understanding: An Inquiry into Human Knowledge Structures*, Lawrence Erlbaum Associates, Publishers (1977).

[16] Soon, W. M., Ng, H. T. and Lim, D. C. Y.: A Machine Learning Approach to Coreference Resolution of Noun Phrases, *Computational Linguistics*, Vol. 27, No. 4, pp. 521–544 (2001).

[17] Szpektor, L. and Dagan, I.: earning Entailment Rules for Unary Templates, *Proceedings of The 22nd International Conference on Computational Linguistics (COLING)*, pp. 849–856 (2008).

[18] Torisawa, K.: Acquiring inference rules with temporal constraints by using Japanese coordinated sentences and noun-verb co-occurrences, *Proceedings of Human Language Technology Conference/North American chapter of the Association for Computational Linguistics annual meeting (HLT-NAACL06)*, pp. 57–64 (2006).

[19] Vapnik, V. N.: *Statistical Learning Theory*, Adaptive and Learning Systems for Signal Processing Communications, and control, John Wiley & Sons (1998).

[20] Walker, M., Cote, S. and Iida, M.: Japanese Discourse and the Process of Centering, *Computational Linguistics*, Vol. 20, No. 2, pp. 193–232 (1994).

[21] Walker, M. A.: Limited attention and discourse structure, *Computational Linguistics*, Vol. 22, No. 2, pp. 255–264 (1996).

[22] Yang, X., Su, J., Lang, J., Tan, C. L., Liu, T. and Li, S.: An Entity-Mention Model for Coreference Resolution with Inductive Logic Programming, *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-08: HLT)*, pp. 843–851 (2008).

[23] 関和広, 藤井敦, 石川徹也: 確率モデルに基づく日本語ゼロ代名詞の照応解消, 言語処理学会第 7 回年次大会発表論文集, pp. 510–513 (2001).

[24] 鳥澤健太郎: 「常識的」推論規則のコーパスからの自動抽出, 言語処理学会第 9 回年次大会発表論文集, pp. 318–321 (2003).

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

[26] 飯田龍, 小町守, 乾健太郎, 松本裕治: NAIST テキストコーパス: 述語項構造と共参照関係のアノテーション, 情報処理学会自然言語処理研究会予稿集, NL-177, pp. 71–78 (2007).