Title	教師なし学習で推定した確率分布によるゼロ代名詞の解析
Author(s)	杉原，大悟
Citation	
Issue Date	2003-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1654
Rights	
Description	Supervisor:鳥澤 健太郎，情報科学研究科，修士

修 士 論 文

教師なし学習で推定した確率分布によるゼロ代名詞の解析

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

杉原 大悟

2003 年 3 月

修 士 論 文

教師なし学習で推定した確率分布によるゼロ代名詞の解析

指導教官　鳥澤健太郎 助教授

審査委員主査　鳥澤健太郎 助教授
審査委員　東条敏 教授
審査委員　島津明 教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

110065 杉原 大悟

提出年月: 2003 年 2 月

概 要

本論文では, 教師なし学習で得られた確率分布による日本語ゼロ代名詞の照応解析について述べる. ゼロ代名詞の照応解析は, 機械翻訳や要約など, 他の自然言語処理アプリケーションを作成する際に非常に有益である. 先行のゼロ代名詞照応解析の研究 [4][5][8] では, 照応解析に用いる情報を, 大規模コーパスと人手によるシソーラスの双方から取得している. 本研究では, 教師無し学習手法のみで推定された確率分布によるゼロ代名詞照応解析手法を提案する. 本研究で扱うゼロ代名詞照応解析のための言語的な資源は 2 つある. 1 めは Rooth[1] の教師無し手法で取得された「名詞の動詞に対する係りやすさ」を表す確率分布である. 実際には, 本研究では, Rooth の手法を日本語に対して拡張した鳥澤の手法 [3] による確率分布を用いる. そして, 2 つめは, Rooth の手法による確率分布を利用して本研究にて推定される「名詞と動詞との位置関係による名詞のゼロ代名詞の先行詞になりやすさ」を表す確率分布である. 本研究では, テキスト中の名詞がゼロ代名詞になる時を表す確率モデルを構築し, それに対して EM アルゴリズム [11] の一般的な手法を適用することで,「名詞の動詞に対する係りやすさ」を表す確率分布から「名詞と動詞との位置関係による名詞の先行詞になりやすさ」を表す確率分布を推定する方法を導出した. 読売新聞 3120 記事を学習データとして確率分布の推定を行い, 推定した確率分布を用いてゼロ代名詞の照応解析実験を行った. Rooth の手法による「名詞の動詞に対する係りやすさを」表す確率分布のみで照応解析を行う場合の精度と, Rooth の手法による「名詞の動詞に対する係りやすさ」を表す確率分布と「名詞と動詞との位置関係による名詞の先行詞になりやすさ」を表す確率分布の両方を用いて照応解析を行う本研究手法の精度を比較したところ, 本研究手法による精度が, Rooth の手法による確率分布のみを用いた場合の精度よりも 11% 高くなることが分かった. これにより, 教師無し学習結果を用いてさらに教師無し学習することにより, ゼロ代名詞照応解析の精度が向上する見通しが得られたことになった. 以上から, 本研究では, 本研究の手法によって取得された確率分布により, ゼロ代名詞照応解析の精度の向上が確かめられたと結論づける.

目次

第1章　はじめに **1**

第2章　関連研究 **3**

2.1　関らによる確率モデルを用いたゼロ代名詞照応解析 . . . 3
2.1.1　関らが扱うゼロ代名詞の種類 . . . 3
2.1.2　関らのシステムの概要 . . . 4
2.1.3　関らが提案する確率モデル . . . 4
2.1.4　関らの行った実験 . . . 6
2.2　河原らによる大規模コーパスから自動取得された格フレーム辞書によるゼロ代名詞照応解析 . . . 7
2.2.1　格フレームの自動構築 . . . 7
2.2.2　河原が扱うゼロ代名詞の種類 . . . 8
2.2.3　河原らのシステムの概要 . . . 8
2.2.4　河原らが行った実験 . . . 9
2.3　教師無し学習にまつわる関連研究 . . . 10

第3章　確率モデル **12**

3.1　動詞はどのような名詞を伴いやすいか . . . 12
3.2　動詞と名詞の「位置関係」 . . . 13
3.3　確率モデル . . . 15
3.3.1　確率モデルの前提 . . . 15
3.3.2　基本確率モデル . . . 16
3.3.3　基本確率モデルの拡張 . . . 18
3.4　Q関数の導出 . . . 21
3.5　パラメータの更新式の導出 . . . 25

第4章　実験 **35**

4.1　Rの抽出 . . . 35
4.2　学習 . . . 37
4.3　正解コーパスの作成 . . . 38
4.3.1　ゼロ代名詞のタグをつけたコーパス . . . 38

4.3.2 ゼロ代名詞タグ付け結果 41
4.3.3 ゼロ代名詞タグ付コーパスの正解の詳細 42
4.4 照応解析実験 44
4.4.1 先行詞を決定するスコアの算出法 45
4.4.2 精度の算出方法 45
4.4.3 意味クラスの粒度 46
4.4.4 実験結果 46

第 5 章 考察と今後の課題 **49**
5.1 考察 49
5.1.1 関らの手法との比較 49
5.1.2 河原らの手法との比較 50
5.1.3 本研究の現時点での結論 50
5.2 今後の課題 51

第1章　はじめに

自然言語による情報の伝達の過程においては, 情報を受け取る側が容易に推測可能な事柄は省略されたり, 代名詞で置き換えられることが多々ある. ある言語表現が, 後に現れる言語表現と同じ内容や対象を指すとき, これらの表現は照応関係 (anaphoric relation) にあるという. この時前者の表現を先行詞 (antecedent), 後者の表現を照応詞 (anaphor) という. 下の例では「次郎」が先行詞になり,「彼」が照応詞になる.「彼」は照応詞として先行詞である「次郎」を参照する.

- 次郎は本屋に行きました.
- 彼は雑誌を二冊買いました.

日本語では代名詞による照応関係の他に, ゼロ代名詞による照応関係も生じる. ゼロ代名詞とは, 省略され見えなくなってしまう照応詞を言い表す言葉である.

- 次郎は本屋に行きました.(0 ガ) 雑誌を二冊買いました.

この場合では, 先行詞「次郎」であり照応詞「0」である.「0」は照応詞として先行詞である「次郎」を参照している. そして, 照応詞「0」はガ格で「買いました」に係っている. 本来「0」はテキストの上に現れない. 実際には下のように見える.

- 次郎は本屋に行きました. 雑誌を二冊買いました.

上記の例のように, 先行詞がすでに文章中にあって, それを参照する形で照応詞が現れる場合を前方照応 (anaphora) という. 逆に, 照応詞が参照する表現が, 照応詞よりも後ろにある場合を後方照応 (cataphora) という. さらには, 照応詞が参照する表現が文章中にはない場合を外界照応 (exophora) という.

ゼロ代名詞は日本語では一般的な現象である. この省略された対象と用言との対応を計算機で求めてやることは, 自動要約や機械翻訳システムなど, 自然言語を扱うアプリケーションを実現するにあたり有益である. 特に日英機械翻訳では, 日本語の主格が省略されやすいのに対し, 英語では主格は必須の要素となるため, ゼロ代名詞照応解析は重要な問題となる. しかし, ゼロ代名詞がテキスト中に現れたことを識別し, そのゼロ代名詞がテキストのどの表現を参照するかは, 人間にとっては容易なことが多いが, 計算機では困難である. 人間は, 自然言語のテキストに長い間接し, ゼロ代名詞を読み抜く知識を蓄積してい

る. 計算機でゼロ代名詞の照応関係を解釈するには, 自然言語についての様々な情報や背景知識を計算機に与える必要がある. ゼロ代名詞を発見するための情報や知識をどのように取得するかについて, 先行のゼロ代名詞照応解析研究においても様々な手法が提案されている. センタリング理論を用いる手法 [12][13] やゼロ代名詞解析のための解析規則を人手で作成する手法 [6] などがある. これらの手法は現在までに多くの成果をあげているが, センタリング理論はシンプルな理論であり, 取り扱える問題が限られている. 解析規則を人手で作成するには多大なコストがかかるという問題がある. 最近では, 大規模コーパスからなんらかの言語的資源を自動的に取得し, 人手で作成した知識を組み合わせてゼロ代名詞の照応解析を行う研究 [4][5] [8][9] が盛んになってきた.

しかし, これらの大規模コーパスを利用する先行研究にしても, なんらかの人手によるコストが介在している. そこで, 本研究では, 教師無し学習手法のみで推定された確率分布によるゼロ代名詞照応解析手法を提案する. 本研究で扱うゼロ代名詞照応解析のための言語的な資源は 2 つある. 1 めは Rooth の教師無し手法 [1] で取得された「名詞の動詞に対する係りやすさ」を表す確率分布である. 実際には, 本研究では, Rooth の手法を日本語に対して拡張した鳥澤の手法 [3] による確率分布を用いる. そして, 2 つめは, Rooth の手法による確率分布を利用して本研究にて推定される「名詞と動詞との位置関係による名詞のゼロ代名詞の先行詞になりやすさ」を表す確率分布である. システムが文中のゼロ代名詞を認識し, その省略されてしまった名詞の参照先である先行詞を推定するために, 本研究では先行詞のゼロ代名詞になりやすさを示す指標として確率の値を用い, この値の大小を比較することで先行詞の推定を行う.

本論文の構成であるが, 次章で本研究と関連のある研究について解説し, 3 章で本研究で用いる確率モデルと確率モデルに対する EM アルゴリズム [11] の適用について述べる.4 章では確率分布を取得するための学習と, 得られた確率分布を用いて行ったゼロ代名詞照応解析実験について述べる.5 章で実験結果を考察し, 今後の展望を述べる.

なお, 本研究が対象とするゼロ代名詞は文内照応を含める前方照応であり, 後方照応と外界照応は対象外とする.

第2章　関連研究

ゼロ代名詞照応解析についての関連研究

ゼロ代名詞の照応解析の研究はこれまでに様々な方法が提案されている. 人手でゼロ代名詞照応解析のための規則を作成し, それをゼロ代名詞照応解析に用いる方法 [6] やセンタリング理論 [12] を用いる方法 [13] 等は盛んに研究されてきた. しかし, 同時にこれらの方法に対しての問題点も議論されてきた. 人手で規則を作成することは非常にコストのかかる作業である. また, センタリング理論はシンプルな理論であり, それを照応解析に応用できる場合は限られている. 最近では大規模コーパスからなんらかの言語的資源を取得し, それをゼロ代名詞照応解析に用いようという研究 [4][5][8] が盛んである. 本研究も, タグ無しコーパスから確率分布を取得してゼロ代名詞照応解析しようという研究であり, その流れの上にあるといってよい. 本節では, 本研究と関連のある最近のゼロ代名詞照応解析研究を紹介する.

2.1　関らによる確率モデルを用いたゼロ代名詞照応解析

関ら [4][5] はタグありタグ無しコーパス双方からそれぞれ確率分布を取得し, それを用いてゼロ代名詞の照応解析を行っている. タグなしコーパスを用いている点, 確率分布を用いている点が本研究と類似している. また, ゼロ代名詞の先行詞を同定するだけでなく, ゼロ代名詞を検出することも併せて行うことを目指している. 本研究と異なる点は, タグ無しコーパスを用いて確率分布を学習する際に人手によるシソーラスを使っている点と, タグありコーパスを用いて学習と評価を同時に行っている点である. 関らは, コーパス中のゼロ代名詞の位置とゼロ代名詞の格が既知であるという前提での先行詞同定実験で, 50.7% の精度で正しいゼロ代名詞の先行詞を出力することができたと報告している.

2.1.1　関らが扱うゼロ代名詞の種類

関らが対象としているゼロ代名詞の種類は, 文内照応を含むガ格, ヲ格, ニ格の前方照応である. 後方照応や外界照応は扱っていない. 本研究も関らに準じている. また, 提題助詞「は」はガ格として扱われ, 関らの研究ではガガ構文を扱うために動詞に生じた２回のガ格を動詞に近い方から, ガ格１, ガ格２のように区別をしているが, 提題助詞「は」にもこ

の規則が適用されている.

2.1.2　関らのシステムの概要

関らのゼロ代名詞検出およびゼロ代名詞補完システムは以下のような流れで処理を行うことを目指している.

1 テキストを JUMAN で形態素解析し, ついで KNP で構文解析する

2 構文解析された結果と IPAL の動詞辞書を参考を用いて, 動詞の充足されていない格をゼロ代名詞の候補として検出する

3 先行詞候補を確率値で評価し, 先行詞を決定する

2 のゼロ代名詞の検出に, 関らは IPAL の動詞辞書を利用している. しかし, IPAL 動詞辞書に登録されている動詞は 861 しかない. 関らは, 未登録の動詞に対しては, IPAL 動詞辞書記載の類義語情報を用いて, 意味的に類議している動詞の格フレームは同じものであると見なして処理を行っている. また, 同じ表記の動詞が複数の格フレームを持つ場合に対して, 関らは同じ動詞の複数の格フレームをマージして使用することでゼロ代名詞の誤検出を防ごうとしている.IPAL 動詞辞書で格フレームを決定できない動詞に対してはガ格のみを必須格としている.

3 のゼロ代名詞補完の時に用いる確率分布として, 関らは,「動詞 v の充足されていない格 c がゼロ代名詞である確率」$P_{zero}(c|v)$ と「動詞 v の格 c に対応するゼロ代名詞 ϕ_c が真のゼロ代名詞である確率」$P(a_i|\phi_c)$ を用いている. 関らはこの 2 つの確率から, 以下の式の値を計算する.

$$P(a_i|\phi_c) \times P_{zero}(c|v) \tag{1}$$

関らは, この値を用いて 2 で検出されたゼロ代名詞候補をさらに絞り込む処理を行っている. 関らは以下の確信度という値を定義している.$P_i(\phi_c)$ とは, あるゼロ代名詞の先行詞候補のうち i 番目に高い (1) の値である.

$$C(\phi_c) = tP_1(\phi_c) + (1-t)(P_1(\phi_c) - P_2(\phi_c))$$

この確信度の意図するところは, 先行詞候補の中で, 他の先行詞候補よりも飛び抜けてゼロ代名詞になりそうなものだけをゼロ代名詞として検出しようということである.

2.1.3　関らが提案する確率モデル

関らは, 確率分布のために, ゼロ代名詞の先行詞候補 a_i とゼロ代名詞 ϕ_c の属性として 6 つの属性を利用している. 関らは 6 つの属性を「ゼロ代名詞の属性」と「先行詞の属性」に分けそれぞれ説明を与えている.

ゼロ代名詞の属性

- ゼロ代名詞を持つ動詞 (v)
- ゼロ代名詞の格 (z):ガ格, ヲ格, ニ格のいずれかをとる

先行詞の属性

- 先行詞候補に現れる助詞 (p_i)
- ゼロ代名詞と先行詞候補間の文間距離 (d_i):ゼロ代名詞と先行詞が同一文内にある場合, すなわち文内照応の場合は $d_i = 0$ となり,n 文前ならば $d_i = n(n > 0)$ となる
- 先行詞候補が連体修飾節に含まれるかどうか (m_i)
- 先行詞候補の意味分類 (n_i):日本語分類語彙表シソーラスに定義されている 544 の意味クラス

$P_{zero}(c|v)$ **の推定**

関らは, ある動詞 v の格 c のゼロ代名詞を, 省略されてしまった動詞の必須格であるとし, ある動詞 v の格 c にゼロ代名詞が現れるかどうかは, その格 c がその動詞の必須格かどうかに依存していると主張している. そして, 動詞 v と格 c が伴って現れる確率が高い程, 格 c は必須格である確率が高く, ある動詞 v の格 c でゼロ代名詞が生じる確率 $P_{zero}(c|v)$ も高くなるとしている.$P_{zero}(c|v)$ は動詞と助詞との共起頻度から以下に定義される.$F(x)$ とは x という事象がコーパス中に起った回数である.

$$P_{zero}(c|v) = \begin{cases} \frac{F(c,v)}{F(v)}, & \text{c がヲ格, ニ格} \\ 1, & \text{c がガ 1 格} \\ \frac{F(\text{ガ 1 格, ガ 2 格},v)}{F(\text{ガ 1 格},v)}, & \text{c がガ 2 格} \end{cases}$$

これらの確率分布は動詞と助詞の共起頻度をタグ無しコーパスからカウントし, 推定される.

$P(a_i|\phi_c)$ **の推定**

関らは $P(a_i|\phi_c)$ を以下に表現している.

$$\begin{aligned} P(a_i|\phi_c) &= P(p_i, d_i, r_i, n_i|v, c) \\ &\approx P(p_i, n_i|v, c)P(d_i)P(r_i) \\ &= P(p_i|n_i, v, c)P(n_i|v, c)P(d_i)P(r_i) \\ &\approx P(p_i|c)P(n_i|v, c)P(d_i)P(r_i) \end{aligned}$$

関らは $P(p_i|c), P(d_i), P(r_i)$ を統語モデルと呼び,$P(n_i|v,c)$ を意味モデルと呼んでいる. そして統語モデル, 意味モデルを関らは以下の式で与えている.

$$\begin{aligned} P(p_i|c) &= \frac{F(p_i,c)}{\sum_j F(p_j,c)} \\ P(d_i) &= \frac{F(d_i)}{\sum_j F(d_j)} \\ P(r_i) &= \frac{F(r_i)}{\sum_j F(r_j)} \\ P(n_i|v,c) &= \frac{F(n_i,v,c)}{\sum_j F(n_j,v,c)} \end{aligned}$$

関らは, 統語モデルのためにゼロ代名詞のタグ付コーパスを人手で作成し, コーパスから各属性ごとの頻度をカウントし, 確率分布の推定を行っている. そして意味モデルについては, タグ無しコーパスから動詞と助詞と名詞の共起頻度をカウントすることで推定している. 名詞は分類語彙表によって意味クラスに汎化される. 分類語彙表に登録されていない名詞は名詞の表記そのものを意味クラスとしている.

本研究の確率分布と関らの確率分布との大きな違いは, 本研究が EM アルゴリズムによってタグ無しコーパスから確率分布を推定するのに対して, 関らはコーパスからカウントした頻度から確率分布を推定している点である. 特に統語モデルはゼロ代名詞のタグ付コーパスから直接にゼロ代名詞が起った場合を観測し, 確率分布を推定している.

2.1.4 関らの行った実験

関らは, 以下の二種類の実験を行っている.

1 ゼロ代名詞の先行詞特定の実験：ゼロ代名詞の位置と格が正しく検出されたという前提で, その動詞の先行詞を出力させて $P(a_i|\phi_c)$ の性能を検証する実験

2 ゼロ代名詞検出を行う実験：$P(a_i|\phi_c)P_{zero}(c|v)$ の性能を検証

関らは,1995 年度の毎日新聞の政治やスポーツなどの報道記事から無作為に 30 記事を選び, これに人手でゼロ代名詞に関するタグを付加した. そして,1 つ抜きの交差確認法により統語モデルの学習と意味モデルと統語モデルを併せたシステムの評価を行っている. つまり,1 記事をテストデータとし, 残り 29 記事で学習にする試行を 30 回行い, その結果を平均している.

1 の実験では 50.7% の精度で正しいゼロ代名詞の先行詞候補を出力することができたと報告している. また, 学習するコーパスの量を変化させて実験し, 少量のタグありコーパスの利用がゼロ代名詞の先行詞同定に有効であると主張している.

2 の実験では, 確信度を用いた出力制限を行わない場合, 式 $P(a_i|\phi_c)P_{zero}(c|v)$ によるゼロ代名詞検出の性能は, 再現率は 90.0% であり, 正解率（適合率）は 25.9% であると報告している. そして確信度による出力制限を行った場合は, 再現率と正解率はトレードオフの関係にあり, 再現率が 50% の時の正解率は 54%[1]である.

2.2 河原らによる大規模コーパスから自動取得された格フレーム辞書によるゼロ代名詞照応解析

河原らは, 大規模のタグ無しコーパスから格フレーム辞書を自動的に構築するという研究 [7] を行っており, この格フレーム辞書を用いてゼロ代名詞の照応解析を行っている [8].

2.2.1 格フレームの自動構築

河原らの格フレーム辞書の自動構築法の基本的な方法は, 用言とその直前の格要素の組を単位として用例を収集し, それらをクラスタリングを行うというものである. 河原らによると, 用言の直前の格要素をペアにして考えると, 用言の用法はほとんど一意に決定されるという. 例えば,「荷物を積む」と「経験を積む」の「積む」の用法の差は, 用言の直前の「荷物を」と「経験を」によって決まる. 河原らの格フレーム自動構築過程は以下のようになる.

1 コーパスのテキストを KNP で構文解析する. 信頼できる用言と格要素の組だけを取り出す. 河原らはこれを「用例」を集めると呼ぶ

2 集められた「用例」を「用言の直前の格要素」ごとにまとめる例えば,「車に荷物を積む」と「トラックに荷物を積む」という用例があったならば,{ 車に, トラックに }「荷物を積む」のようにまとめる

3 シソーラスを用いて, 用例パターンのクラスタリングを行い, 用言の用例ごとの格フレームを作る

用例パターンは, 一つの用言について直前の格要素の数だけ存在している. 3 では用法がほとんど同じ用例パターンをマージしている. 例えば「従業員が車に荷物を積む」と「トラックに物資を積む」の「積む」のヲ格の格フレームの用法はほとんど同じであり, これをまとめてしまうということである. 河原らは, 用例パターンのクラスタリングを, 用例パターン間の類似度を用いて行っている. 用例パターンの類似度は, その格の一致度と２つの用例パターンの共通格に現れる格要素の類似度の積で表される. 格の一致度とは, 用例パターン P_1, P_2 に含まれる全ての格要素に対する,P_1, P_2 の共通格に含まれている格要素の割合である. 河原らは, 格要素の類似度の計算に, 日本語語彙大系を用いている.

[1]参考文献 [5] 中に直接の言明はないが, 図から読み取れる

2.2.2 河原が扱うゼロ代名詞の種類

河原らは「ゼロ代名詞」という言葉よりもむしろ「格要素の省略」という言葉を用いている. 河原らは用言 (動詞, 形容詞, 名詞+判定詞) とサ変名詞に対して, 直接係り受け関係を持っている格要素と, 省略されてしまった格要素についてのタグ付けを行っている[2]. タグを付加した格は, ガ格, ニ格, ヲ格である. また, 河原らは省略タグを付与する過程で, 省略の対象が存在するのは, ほとんどが以下の「限られた範囲」にあると報告している.

- 対象用言がある文中
- 前文の重要な要素
- 記事の最初の文の重要な要素

上の「重要な要素」とは, 以下である.

- 主節の用言に係る格要素 (省略された指示対象も含む)
- 主節に係る従属節で, かつ「ので」のようなスコープの広い従属節に含まれる格要素 (省略された指示対象も含む)
- 提題助詞「は」が付属する名詞句中に含まれる語

また, 河原らは動作主体の省略が, 具体的な対象を指さず不特定の人々を指しているような場合には「不特定：人」というタグを付与し, これを省略された指示対象であるとしている. 以上から, 河原が対象としている省略 (ゼロ代名詞) は, 用言だけでなくサ変名詞も含まれている上に, 文内照応を含む前方照応とそれに含めて後方照応[3] と外界照応までを含むことになり, 本研究や関らが扱っているゼロ代名詞の範囲よりもより広い領域をカバーしていると言える.

2.2.3 河原らのシステムの概要

1 テキストを構文解析し, 格解析する

2 格解析結果から, 各用言の省略されている格要素を認識する

3 省略の指示対象を推定する

[2]河原らは, 省略の指示対象として, 文節内の自立語が単位としている. しかし, 必要によっては単語を単位とすると報告している.

[3]同じ文中の後方照応.

河原らによると1の格解析は, テキストの用言に対して, 用言の用法ごとに用意されている格フレームの中から, 入力文に最も合うものを選択し, 同時に入力文の格要素との対応づけを行う処理であるという. そして,2では, 格フレームに入力文の格要素と対応つけられていない格があり, それがガ格, ヲ格, ニ格であれば, その格が省略された格であると認識する. 関らの手法でのゼロ代名詞検出の時には,IPAL 動詞辞書の動詞の格フレーム辞書とテキスト中の動詞の格要素を比較するが, この時動詞の多義性は考慮されていない. これに対して河原らは自動で用言の用例ごとの格フレーム辞書を構築することで, 動詞の多義性に対処している.

河原らは, 省略されたと認識した格の指示対象の推定を行う時に, その格フレームの用例群を用いている. まず, 先に述べた範囲から指示対象の候補を選びだし, その候補に対してスコア付を行う. スコアが最大で, かつ閾値を超えていれば指示対象として決定するという. スコアが閾値を超えないようならばガ格とニ格は「不特定：人」を, ヲ格は「不特定：もの」として推定する. スコアとは, 格フレームの用例群と候補間の類似度であり, 以下の式で計算している.

$$\text{スコア} = \text{重み} \times \text{類似度}$$

重みと類似度として河原らは以下に定めている.

重み：テキストの中の体言で, 重要な要素でないもの, または処理中の用言よりも後ろにあるものは0.6, それ以外1.0

類似度：省略候補と格フレームの省略格の格用例群との類似度で, NTT の日本語語彙体系を用いて計算する

2.2.4 河原らが行った実験

河原らは先に作った省略のタグのついたコーパスに対して, システムの性能評価の実験を行っている. そして98.6% の省略の指示対象が上記の「限られた範囲」内にあったと報告している. また, 河原らはシステムの適合率と再現率は以下になったと報告している. 総じて, 関らの手法よりも性能が高い. 動詞の多義性を考慮することによって再現率の点で関らを上回り, 用言の複数の格要素を同時に考慮することにより適合率の点でも関らを上回ったものと考えられる.

	適合率	再現率
用言	175/259(68.7%)	178/251(70.9%)
サ変名詞	64/154(41.6%)	64/118(54.2%)
計	242/413(58.6%)	242/369(65.6%)

河原の照応解析実験結果

2.3 教師無し学習にまつわる関連研究

本研究では,「動詞と名詞が何格で関わりやすいか」という確率分布を Rooth[1] の教師なし手法で取得し, これをゼロ代名詞照応解析に用いている. 実際には, 本研究では, Rooth の手法を日本語に対して拡張した鳥澤の手法 [3] による確率分布を用いる. 本節ではこの二つの研究について解説する.

Rooth, 鳥澤の両手法ともに, ベースになるのは以下のモデルである.$\langle v, c\rangle$ は, 動詞と名詞と動詞の関係の組であり,n は名詞を表す.

$$\begin{aligned} P(\langle v,c\rangle, n) &= \sum_{a\in Class} P(\langle v,c,\rangle, n, a) \\ &= \sum_{a\in Class} P(\langle v,c\rangle|a)P(n|a)P(a) \end{aligned}$$

a は動詞や名詞が属するクラスである. このクラス a は学習コーパスには現れない.EM アルゴリズムに基づく教師なし学習法で学習される過程において, 単語の意味クラスとして生成されることになる. 両手法ともに,$P(\langle v,c\rangle, n)$ を, 意味クラスの生じる確率 $P(a)$, 意味クラスを条件部とする条件付確率 $P(n|a)$, $P(\langle v,c\rangle|a)$ によって表現することで, データスパースネスに対処している.

動詞の c は,Rooth のモデルでは,「(.as:s) 自動詞の主語」, 「(.aso:s) 他動詞の主語」,「(.aso:o) 他動詞の目的語」の 3 種類をとる. 対して, 日本語を扱う鳥澤のモデルだと,c は日本語の助詞と「助詞がない」ことを表す記号になっている[4].

Rooth, 鳥澤の両手法共に, 求めようとしているのは $P(\langle v,c\rangle, n)$ を構成する $P(a)$,$P(n|a)$, そして $P(\langle v,c\rangle|a)$ である. これらの確率値をパラメータとして, EM アルゴリズムの手法を用いて計算している.

両手法における確率分布の推定法

学習データとしてタグ無しコーパスから「動詞 v」と「名詞 n」と「動詞と名詞との間の関係 c」を取得し, その個々のデータ全体を L, 個々のデータを $y_i = \langle v_i, c_i, n_i\rangle$ とする.v と n と c の組がコーパス中に m 個あったとすると, 全データは以下のように表される.

$$L = \langle\langle v_1, c_1, n_1\rangle, ..., \langle v_m, c_m, n_m\rangle\rangle$$

ここで, コーパス上で観測できない隠れ変数として, クラス $a \in Class$ を考える. 個々の観測データが現れたときにクラス a が生じる確率は, 以下になる.

$$P(a, y_i) = p(\langle v_i, c_i\rangle, n_i, a) = P(\langle v_i, c_i\rangle|a)P(n_i|a)P(a)$$

パラメータの集合を θ とする.EM アルゴリズムでは, 与えられた全データの対数尤度をパラメータの関数と見なし, これを最大化するようなパラメータを繰り返し探していく. Q 関数は下になると思われる.

[4]「30 日東京に行く」の「30 日」のように, 助詞なしで動詞との間に関係を生じさせている場合を表す

$$Q(\hat{\theta},\theta)=\sum_{L}\sum_{a\in Class}P(a|\langle v,c\rangle,n)logP_{\theta}(\langle v,c\rangle|a)P(n|a)P(a)$$

両手法のモデルともに文脈を考慮していないので, 全データの対数尤度を最大化するようなパラメータを推定する式はシンプルに以下のような形になる. 両手法を記した論文 [1],[3] に直接の記述はないが, 制約条件 $\sum_{a\in Class}P(a)=1$ の元に Q 関数を最大化するようなパラメータを求めたものと思われる.

$$\begin{aligned}
P_{\hat{\theta}}(a) &= \frac{\sum_{\langle v_i,c_i,n_i\rangle\in L}P_{\theta}(a|\langle v_i,c_i,n_i\rangle)}{|L|}\\
P_{\hat{\theta}}(\langle v,c\rangle|a) &= \frac{\sum_{\langle v,c,n_i\rangle\in L}P_{\theta}(a|\langle v,c,n_i\rangle)}{\sum_{\langle v_i,c_i,n_i\rangle\in L}P_{\theta}(a|\langle v_i,c_i,n_i\rangle)}\\
P_{\hat{\theta}}(n|a) &= \frac{\sum_{\langle v_i,c_i,n\rangle\in L}P_{\theta}(a|\langle v_i,c_i,n\rangle)}{\sum_{\langle v_i,c_i,n_i\rangle\in L}P_{\theta}(a|\langle v_i,c_i,n_i\rangle)}
\end{aligned}$$

Rooth, 鳥澤両手法におけるパラメータ推定式

パラメータの更新式の形は Rooth, 鳥澤共に同じ形をしている. 両手法ともに, 結果として, 得られた確率分布は, クラス a は単語の意味クラスと格フレームとしての働きをすることが報告されている. また, 両手法ともに, 名詞や動詞は複数のクラスに同時に関わることができるソフトクラスタリングの手法であり, 名詞や動詞の多義性を扱うことができる. 鳥澤は,「早大」という名詞を「大学」と「スポーツチーム」の二つの意味で扱えていると報告している. 鳥澤のモデルは Rooth のモデルに日本語としての意味を与え, さらなる拡張を加えている.$P(v,c,n)$ は「名詞 n が助詞 c を介して動詞 v に係っている確率」と解釈できるとしている. また, 鳥澤の手法では, 構文解析結果を用いて学習データに重み付けをしている. その構文解析器の学習には EDR コーパスが用いられているが, EDR コーパスにおいて人手で与えられている正解は, 鳥澤の手法で推定される確率分布の正解にはなりえないので, 本研究では鳥澤の手法を教師無し手法として見なす. 本研究は日本語ゼロ代名詞の照応解析を目的としているので, この解釈を元に確率分布を利用する.

第3章　確率モデル

本研究では, 文章に現れる名詞のゼロ代名詞になりやすさを確率の値で評価し, 照応解析を行う. 名詞のゼロ代名詞としての出現しやすさを考える上で, 本研究ではおおまかに, 以下の二つの側面があると考える.

(1) 動詞はどのような名詞を伴いやすいか

(2) 名詞と動詞の「位置関係」

本章では, これらの側面からテキスト中の名詞がゼロ代名詞になる時を表す確率モデルを構築し, そこから「名詞と動詞の位置関係」に関する確率分布を推定するための方法を導出する.

3.1　動詞はどのような名詞を伴いやすいか

動詞が述語として文を形成するに当たって, 重要な概念に「格 (case)」というものがある. 動詞には, 述語として文を形成するにあたり, 自らの表す動きや状態を作るためにどのような名詞句をとるかが基本的に決まっている, という現象が存在する. この時, 以下のような言葉を導入してこの現象を説明する.

- 格支配：用言が自らが帯びる語彙的な性質に応じて, 文の形成に必要な名詞句を選択的に要求する働き
- 格要素：用言の格支配を受けて文を作る名詞句のこと
- 格：用言の格支配を受ける名詞句の, 用言に対する関係のこと

  例文：達郎がどら焼きを食べた.

この例だと, 動詞「食べた」が「達郎が」と「どら焼きを」という名詞句を要求する働きが格支配であり, これら名詞句が格要素である. そして「食べた」という動作と「誰が食べた」という関係を持つのが「達郎」であり,「食べた」の「何を食べた」という関係を持つのが「どら焼き」である. このような名詞句の用言に対する関係が「格」と呼ばれるものである.

一般に,「食べる」という動詞のガ格には,「人」や「動物」に属するであろう名詞のほうが,「乗り物」や「時間」などに属するであろう名詞より選ばれやすい. このことは仮に「達郎が」がゼロ代名詞になり見えなくなってしまった後でも, この文の読み手にとっては変らないのである. つまり,「食べる」のガ格のゼロ代名詞には, 「乗り物」や「時間」などに属するであろう名詞よりも, 「人」や「動物」に属するであろう名詞のほうがなりやすいと言えるのである.

このような「動詞はどんな名詞を伴いやすいか」という情報はとして, 本研究ではRooth[1]の手法でEMアルゴリズムによってタグ無しコーパスから推定された確率分布を用いる.

$$P(v,c,n) = \sum_{a \in Class} P(\langle v,c \rangle | a) P(n|a) P(a)$$

実際には, 本研究では, Roothの手法を日本語に対して拡張した鳥澤の手法[3]による確率分布を用いる. この式の意味するところは, 鳥澤によると「名詞nと動詞vに格cが生じている確率」である. 本研究では, 動詞vに対する格cのゼロ代名詞の先行詞を探すことを, 動詞vと格cの関係を持ちやすいテキスト中の名詞を探すことによって行う. また, この確率分布を用いて本研究では, 「名詞と動詞の位置関係」に関する確率分布の推定を行う.

3.2 動詞と名詞の「位置関係」

本研究では, テキスト中の名詞のゼロ代名詞の先行詞になりやすさは, 名詞と動詞との間の意味的な関係だけでなく, 名詞と動詞との距離, 名詞に後置する助詞など, 様々な統語的な要素によっても左右されると考える. 動詞に近い名詞のほうが, 動詞に遠い名詞よりもゼロ代名詞になりやすいといった事柄である. 本研究では, そのような「名詞と動詞との位置関係」は名詞と動詞との距離だけではなく, 文章における名詞の位置や付属する助詞など, 様々な素性から成り立っていると考える.Rを名詞と動詞との位置関係を表すベクトルとすると, 本研究ではRは以下のように考える.

$$R = \langle d, c1, rentai, head, eachid \rangle \tag{3.1}$$

1 d:名詞と動詞の文間距離. $d \in \{0, 1, 2, ...\}$

2 c1:名詞についている助詞. $c1 \in \{ga, wo, ni, kara, ...\}$

3 rentai:名詞が連体修飾節に含まれているか. $rentai \in \{0, 1\}$

4 head:名詞がテキストの先頭の段落に含まれているかどうか. $head \in \{0, 1\}$

5 eachid:複数の名詞とある動詞が同じ$\langle d, c1, rentai, head \rangle$を持つ時, それらを区別するid. $eachid \in \{1, 2, ...\}$

このRの値は, タグ無しコーパスから観測可能である. しかし, 本研究では, この観測可能な素性に加えて, ゼロ代名詞に関する隠し素性として, 素性 $c2$ を考える. この隠し素性 $c2$ は, コーパスからは観測不能であり, 他のRの素性とは区別して考える必要があるからである.

c2:名詞が動詞に何格で係るか. $c2 \in \{ga, wo, ni\}$

本研究では, この R と $c2$ の値ごとに名詞のゼロ代名詞になりやすさは変化すると考える. 以下に素性についての説明を与える. なお, 本研究は前方照応のみを扱うので, 動詞に対する名詞とは, 動詞の前に出現した名詞のみを対象としている.

d:文間距離

名詞と動詞との文間の距離とは, 動詞を含む文を基準として, 動詞と名詞との間の文がどれだけ離れているかを表す. d は0から正の整数で表され, $d = 0$ の場合は文内照応に相当する. 一般的に, ゼロ代名詞は先行詞から遠ければ遠いほど現れにくくなると言われている. 同じ対象を指し示す名詞の出現が近ければ近いほど, 後に出てくるほうの名詞は省略されやすくなる. この素性については関らの研究 [4] を参考にした.

c1:名詞についている助詞

名詞に後置する助詞の種類によって, その名詞のゼロ代名詞のなりやすさは変化すると思われる. 例えば, 提題助詞「は」がついている名詞は, その文章の主題を表すことが多く, そのような名詞は文を超え, 先行詞から離れた場所のゼロ代名詞として出現することが考えられる. これらの名詞につく助詞は, それぞれの助詞に与えられた番号で表される.

rentai:名詞が連体節に含まれているかどうか

名詞を修飾する動詞に係る名詞は, 連体修飾句に含まれているとしてこれを区別する. 連体修飾句に含まれている名詞は, 被修飾名詞の内容を表現するために特別に現れたものであると考えられるからである. よって, 文章全体を通しては現れにくく, ゼロ代名詞にもなりにくいと考える. これは1,0の値で, 名詞が連体修飾句に含まれているか, そうでないかを表す.

head:名詞が先頭の段落に含まれているかいないか

新聞記事などでは特にそうであるが, 文章の最初の段落では, 文章全体を述べる表現や要約に費やされることが多い. 最初の段落に含まれる表現は, その文章全体を通して使われる表現であり, ゼロ代名詞になりやすいと考えられる. これは1,0の値で, 名詞が最初の段落に含まれているか, そうでないかを表す.

eachid:複数の名詞と1つの動詞が同じRを持った時, それらを区別するためのid

$\langle d, c1, rentai, head \rangle$ はいずれもタグ無しコーパス上で観測可能である．$eachid$ はある動詞の前の複数の名詞に現れる同じ $\langle d,c1,rentai,head \rangle$ を区別する番号である．これは, 同じ R の値を持つ名詞に対して与えられた文間距離とは違った意味での距離である. すなわち, 同じパターンの「位置関係」を持った名詞のどちらかがよりゼロ代名詞になりやすいかを表す.

隠し素性c2:ゼロ代名詞が動詞に何格で係るか

この素性はゼロ代名詞が何格で現れ易いかを表す．「ゼロ代名詞はニ格で現れやすい」などといった状況を表現する．この素性は $c1$ の名詞に後置する助詞素性と併せて,「ある助詞の名詞は何格でゼロ代名詞になりやすい」という状況を表現できる．例えば, 提題助詞「は」を伴った名詞は「ガ格」でゼロ代名詞になりやすい, という状況を表す. この素性は他の素性とは異なり, 素の文を表面から見ただけでは観測できない. また, この素性は, 格を表す番号で表される.

名詞と動詞の位置関係についての確率分布

本研究では, 動詞とある名詞の間に位置関係を R の値によって区別する. そして, ある R の値を持つ名詞がテキスト中に現れた時のゼロ代名詞の先行詞になる度合いを確率分布として表現する.

3.3 確率モデル

本節では, テキスト中の名詞がゼロ代名詞になる時を表す確率モデルを構築し, それに対してEMアルゴリズム[11]の手法を適用し, R に関する確率分布の推定法を導く．以後「名詞と動詞との位置関係」を R と呼び, ある「R を表現するベクトルの値」r_t があったときに,「この値を持つ名詞が他の値を持つ名詞を抑えてゼロ代名詞になる」という確率モデルを考える.

3.3.1 確率モデルの前提

本研究では, 確率モデルを考える上で, 以下のような前提を用意する.

1 コーパス全体に渡って動詞は V 回現れる.

2 「名詞と動詞の位置関係」を考える範囲は, 記事内に限定される. 本研究では前方照応のみを扱う.

3 動詞の前にある R は r_t で区別される. 同じ動詞に同じ R の値が同時には出現しない.

4 $\sum_s P(r_s) = 1.0$ を満たす.

5 動詞の前の名詞のうち, どれか一つが必ずゼロ代名詞となり, その他の名詞はゼロ代名詞にはならないものとする.

6 R の値 $\langle d, c1, rentai, head, eachid \rangle$ はタグ無しコーパス上で観測可能である. これに対して, $c2$ は観測不能であり, これは区別される.

3.3.2 基本確率モデル

まず, 簡単のため, $c2$ を考慮しない, タグ無しコーパス上で観測できる R の値 r_t のみを扱う確率モデルを考える. 名詞が R の値として r_t を持っているとき, その名詞のゼロ代名詞になりやすさを与える確率モデルを考える. ある動詞 v_j とその前に観測される jk 個の名詞と R の集合を x_j とすると,x_j は以下のように表せる.

$$x_j = \langle n_1, ..., n_{jk}, v_j, R_1, ..., R_{jk} \rangle \tag{3.2}$$

ここで,x_j があった時, その中の R に値 r_t を持つものがあり, かつ, その R の名詞が他の名詞を抑えてゼロ代名詞となる確率 $P(x_j, \langle r_t, 1 \rangle)$ を考える. ここで注意しなければならないことは,$P(r_t)$ と $P(\langle r_t, 1 \rangle)$ は全く別の確率であるということである.

いくら $P(r_t)$ が高くとも, それはその r_t をとる名詞がゼロ代名詞になることを示さない. テキストに最も現れやすい名詞と動詞の位置関係の名詞が, 最もゼロ代名詞の先行詞になるわけではないからである. そこで本研究では「名詞がゼロ代名詞の先行詞になる」という状況を数字の 1, 逆に「名詞がゼロ代名詞の先行詞にならない」という状況を 0 で表す.

- $P(r_t)$:R の値 r_t の現れる確率
- $P(\langle r_t, 1 \rangle)$:$R$ の値 r_t が現れ, かつその R の値 r_t を持つ名詞がゼロ代名詞の先行詞になる確率
- $P(\langle r_t, 0 \rangle)$:$R$ の値 r_t が現れ, かつその R の値 r_t を持つ名詞がゼロ代名詞の先行詞にならない確率

もし, r_t の値を持つ R が x_j に現れない場合は, $P(x_j, \langle r_t, 1 \rangle)$ は 0 になる. r_t の値を持つ R が x_j の i 番目の名詞に対応して現れた場合を考える.

$$\begin{aligned} P(x_j, \langle r_t, 1 \rangle) &= P(\langle n_1, ..., n_{jk}, v_j, R_1, ..., R_{jk} \rangle, \langle r_t, 1 \rangle) \\ & \quad R_i \text{ の出現はすなわち } r_t \text{ の出現なので} \\ &= P(n_1, ..., n_{jk}, v_j, R_1, ...R_{i-1}, R_{i+1}, ..., R_{jk}, \langle r_t, 1 \rangle) \end{aligned} \tag{3.3}$$

同様に, r_t 以外の値を持つ R の出現も, R ごとの値の出現と言い換えることができる. k 番目の名詞に対応して出現した R の値を r_{R_k} とすると, 以下のようになる.

$$\begin{aligned} &P(x_j, \langle r_t, 1\rangle) \\ &\quad = P(n_1, ..., n_{jk}, v_j, r_{R_1}, ...r_{R_{i-1}}, r_{R_{i+1}}, ..., r_{R_{jk}}, \langle r_t, 1\rangle) \end{aligned} \tag{3.4}$$

また, 名詞と動詞の出現は R に左右されず, さらにゼロ代名詞とならない i 番目以外の名詞の出現は動詞にも左右されないと仮定すると以下のようになる.

$$= P(n_1)\dots P(n_{i-1})P(n_{i+1})\dots P(n_{jk})P(n_i, v_j)P(r_{R_1}, ...r_{R_{i-1}}, r_{R_{i+1}}, ..., r_{R_{jk}}, \langle r_t, 1\rangle) \tag{3.5}$$

また, r_t の値で R をとる i 番目の名詞以外はゼロ代名詞をにならない. これを数字の 0 で表現すると以下のようになる.

$$\begin{aligned} = \; & P(n_1)\dots P(n_{i-1})P(n_{i+1})\dots P(n_{jk}) \\ & P(n_i, v_j)P(\langle r_{R_1}, 0\rangle, ..., \langle r_{R_{i-1}}, 0\rangle, \langle r_{R_{i+1}}, 0\rangle, ..., \langle r_{R_{jk}}, 0\rangle, \langle r_t, 1\rangle) \end{aligned} \tag{3.6}$$

R の出現は互いに独立であると仮定する.

$$\begin{aligned} &= P(n_1)\dots P(n_{i-1})P(n_{i+1})\dots P(n_{jk})P(n_i, v_j)P(\langle r_{R_1}, 0\rangle)\dots \\ &\quad \dots P(\langle r_{R_{i-1}}, 0\rangle)P(\langle r_{R_{i+1}}, 0\rangle)\dots P(\langle r_{R_{jk}}, 0\rangle)P(\langle r_t, 1\rangle) \end{aligned} \tag{3.7}$$

r_t はコーパス上から観測可能である. よって $P(r_t)$ は既知である. これを確率分布として推定する必要はないので, これを分離する. 分離にはベイズの定理を用いる.

$$\begin{aligned} &= P(n_1)\dots P(n_{i-1})P(n_{i+1})\dots P(n_{jk})P(n_i, v_j)P(r_{R_1})P(0|r_{R_1})\dots \\ &\quad \dots P(r_{R_{i-1}})P(0|r_{R_{i-1}})P(r_{R_{i+1}})P(0|r_{R_{i+1}})\dots P(r_{R_{jk}})P(0|r_{R_{jk}})P(r_t)P(1|r_t) \end{aligned} \tag{3.8}$$

最終的には,

$$\begin{aligned} &P(x_j, \langle r_t, 1\rangle) \\ &\quad = P(n_1)\dots P(n_{i-1})P(n_{i+1})\dots P(n_{jk})P(n_i, v_j)P(r_{R_1})\dots P(r_{R_{jk}}) \\ &\qquad \dots P(0|r_{R_1})\dots P(0|r_{R_{i-1}})P(0|r_{R_{i+1}})\dots P(0|r_{R_{jk}})P(1|r_t) \end{aligned} \tag{3.9}$$

まとめると,

$$P(x_j, \langle r_t, 1\rangle) = \begin{cases} 3.9, & r_t \text{ な } R \text{ が } x_j \text{ に現れ, その } R \text{ は } i \text{ 番目の名詞と対応している時} \\ 0, & r_t \text{ の値を持つ } R \text{ が } x_j \text{ に現れない場合} \end{cases}$$

3.3.3 基本確率モデルの拡張

基本モデルでは, タグ無しコーパス上で観測できる R の値 r_t の素性のみを扱っている. しかし, この基本モデルだけでは, 先行詞とゼロ代名詞との間の格の交換については扱えない. 例えば, 先行詞が「は」という助詞を持っていたときに, その先行詞がゼロ代名詞で現れる場合は「ヲ格」よりも,「ガ格」で現れやすいという現象は, この基本モデルでは扱えないのである. そこで, 基本モデルを拡張して, ゼロ代名詞と動詞との「格」を表す隠し素性も扱えるようにする. 基本確率モデルの時と同様に, ある動詞 v_j とその前に観測される jk 個の名詞と R の集合を x_j とすると,x_j は以下のように表せる.

$$x_j = \langle n_1, ..., n_{jk}, v_j, R_1, ..., R_{jk}\rangle \tag{3.10}$$

また動詞とゼロ代名詞との間に考慮する格の種類を n 個とする. 考慮する全ての格は以下のように表せる.

$$c2 = \langle c_0, ..., c_p, ..., c_n\rangle \tag{3.11}$$

x_j があった時, その中の R に値 r_t を持つものがあり, かつ, その R の名詞が他の名詞を抑えて格 c_p でゼロ代名詞となる確率 $P(x_j, \langle r_t, c_p, 1\rangle)$ を考える. もし,r_t の値を持つ R が x_j に現れない場合は, $P(x_j, \langle r_t, c_p, 1\rangle)$ は 0 になる. r_t の値を持つ R が x_j の i 番目の名詞に対応して現れた場合を考える. これは基本モデルの時と同様である.

$$\begin{aligned} P(x_j, \langle r_t, c_p, 1\rangle) &= P(\langle n_1, ..., n_{jk}, v_j, R_1, ..., R_{jk}\rangle, \langle r_t, c_p, 1\rangle) \\ &\quad R_i \text{ の出現はすなわち } r_t \text{ の出現なので} \\ &= P(n_1, ..., n_{jk}, v_j, R_1, ...R_{i-1}, R_{i+1}, ..., R_{jk}, \langle r_t, c_p, 1\rangle) \end{aligned} \tag{3.12}$$

同様に,r_t 以外の値を持つ R の出現も,R ごとの値の出現と言い換えることができる. k 番目の名詞に対応して出現した R の値を r_{R_k} とすると, 以下のようになる.

$$\begin{aligned} &P(x_j, \langle r_t, c_p, 1\rangle) \\ &\quad = P(n_1, ..., n_{jk}, v_j, r_{R_1}, ...r_{R_{i-1}}, r_{R_{i+1}}, ..., r_{R_{jk}}, \langle r_t, c_p, 1\rangle) \end{aligned} \tag{3.13}$$

また, 名詞と動詞の出現は R に左右されず, さらにゼロ代名詞とならない i 番目以外の名詞の出現は動詞にも左右されないと仮定すると以下のようになる.

$$= P(n_1)\ldots P(n_{i-1})P(n_{i+1})\ldots P(n_{jk})P(n_i, v_j)P(r_{R_1}, ...r_{R_{i-1}}, r_{R_{i+1}}, ..., r_{R_{jk}}, \langle r_t, c_p, 1\rangle) \tag{3.14}$$

そして, i 番目の名詞は格 c_p で動詞と関係を持っているので,

$$P(n_i, v_j) = P(n_i, v_j, c_p) \tag{3.15}$$

となり,3.14 は以下になる.

$$= \quad P(n_1)\ldots P(n_{i-1})P(n_{i+1})\ldots P(n_{jk})P(n_i,v_j,c_p)P(r_{R_1},...r_{R_{i-1}},r_{R_{i+1}},...,r_{R_{jk}},\langle r_t,c_p,1\rangle) \tag{3.16}$$

R の出現は互いに独立であると仮定すると.

$$\begin{aligned} &= P(n_1)\ldots P(n_{i-1})P(n_{i+1})\ldots P(n_{jk})P(n_i,v_j,c_p)P(r_{R_1})\ldots \\ &\quad \ldots P(r_{R_{i-1}})P(r_{R_{i+1}})\ldots P(r_{R_{jk}})P(\langle r_t,c_p,1\rangle) \end{aligned} \tag{3.17}$$

さて, 前提より, 動詞の前の名詞でゼロ代名詞になるのは一つだけである. さらに, 今は r_t の R を持つ先頭から i 番目の名詞のみがゼロ代名詞になる確率を考えているので, i 番目以外の全ての名詞はゼロ代名詞をとらず, i 番目の名詞も格 c_p 以外のゼロ代名詞をとらない. ゼロ代名詞とならないことを数字の 0 で表現すると以下のようになる.

$$\begin{aligned} = \quad & P(n_1)\ldots P(n_{i-1})P(n_{i+1})\ldots P(n_{jk})P(n_i,v_j,c_p)P(r_{R_1},\langle c_0,0\rangle,\ldots\langle c_n,0\rangle)\ldots \\ & \ldots P(r_{R_{i-1}},\langle c_0,0\rangle,\ldots\langle c_n,0\rangle)P(r_{R_{i+1}},\langle c_0,0\rangle,\ldots,\langle c_n,0\rangle)\ldots \\ & \quad \ldots P(r_{R_{jk}},\langle c_0,0\rangle,\ldots\langle c_n,0\rangle)P(r_t,\langle c_0,0\rangle,\ldots,\langle c_p,1\rangle,\ldots,\langle c_n,0\rangle) \end{aligned} \tag{3.18}$$

基本モデルの時と同様に, $P(r_t)$ を分離する.

$$\begin{aligned} = \quad & P(n_1)\ldots P(n_{i-1})P(n_{i+1})\ldots P(n_{jk})P(n_i,v_j,c_p)P(r_{R_1})P(\langle c_0,0\rangle,\ldots\langle c_n,0\rangle|r_{R_1})\ldots \\ & \ldots P(r_{R_{i-1}})P(\langle c_0,0\rangle,\ldots\langle c_n,0\rangle|r_{R_{i-1}})P(r_{R_{i+1}})P(\langle c_0,0\rangle,\ldots,\langle c_n,0\rangle|r_{R_{i+1}})\ldots \\ & \ldots P(r_{R_{jk}})P(\langle c_0,0\rangle,\ldots\langle c_n,0\rangle|r_{R_{jk}})P(r_t)P(\langle c_0,0\rangle,\ldots,\langle c_p,1\rangle,\ldots,\langle c_n,0\rangle|r_t) \end{aligned} \tag{3.19}$$

ここで以下の乗法定理を用いる.

$$\begin{aligned} P(A_0,A_1,...,A_n) &= P(A_0)P(A_1|A_0)P(A_2|A_0,A_1) \\ &\quad \times\ldots\times P(A_n|A_0,A_1,\ldots,A_{n-1}) \end{aligned}$$

上の定理を用いて 3.19 を変形する.

$$
\begin{aligned}
= \;& P(n_1)\dots P(n_{i-1})P(n_{i+1})\dots P(n_{jk})P(n_i,v_j,c_p)P(r_{R_1})P(\langle c_0,0\rangle|r_{R_1})P(\langle c_1,0\rangle|r_{R_1},c_1)\dots \\
& \qquad\qquad\qquad\qquad \dots P(\langle c_n,0\rangle)|r_{R_1},\langle c_0,0\rangle,\dots,\langle c_{n-1},0\rangle)\dots \\
& \dots P(r_{R_{i-1}})P(\langle c_2,0\rangle|r_{R_{i-1}})P(\langle c_1,0\rangle|r_{R_{i-1}},c_0)\dots \\
& \qquad\qquad\qquad\qquad \dots P(\langle c_n,0\rangle)|r_{R_{i-1}},\langle c_0,0\rangle,\dots,\langle c_{n-1},0\rangle) \\
& P(r_{R_{i+1}})P(\langle c_0,0\rangle|r_{R_{i+1}})P(\langle c_1,0\rangle|r_{R_{i+1}},c_0)\dots \\
& \qquad\qquad\qquad\qquad \dots P(\langle c_n,0\rangle)|r_{R_{i+1}},\langle c_0,0\rangle,\dots,\langle c_{n-1},0\rangle)\dots \\
& \dots P(r_{R_{jk}})P(\langle c_0,0\rangle|r_{R_{jk}})P(\langle c_1,0\rangle|r_{R_{jk}},c_0)\dots \\
& \qquad\qquad\qquad\qquad \dots P(\langle c_n,0\rangle)|r_{R_{jk}},\langle c_0,0\rangle,\dots,\langle c_{n-1},0\rangle) \\
& P(r_t)P(\langle c_0,0\rangle|r_t)P(\langle c_1,0\rangle|r_{R_i},c_0)\dots \\
& \qquad\qquad\qquad\qquad \dots P(\langle c_p,1\rangle|r_t,c_0,c_1,...,c_{p-1})\dots \\
& \qquad\qquad\qquad\qquad\qquad \dots P(\langle c_n,0\rangle)|r_t,\langle c_0,0\rangle,\dots,\langle c_{n-1},0\rangle)
\end{aligned} \tag{3.20}
$$

ここで,$\langle c_i,0\rangle$ および $\langle c_i,1\rangle$ はそれぞれの R にのみ依存すると仮定する. この時条件づき確率は, 以下のようになるので,

$$
P(\langle c_n,0\rangle)|r_t,\langle c_0,0\rangle,\dots,\langle c_{n-1},0\rangle) \;=\; P(\langle c_n,0\rangle)|r_t) \tag{3.21}
$$

よって,

$$
\begin{aligned}
= \;& P(n_1)\dots P(n_{i-1})P(n_{i+1})\dots P(n_{jk})P(n_i,v_j,c_p) \\
& P(r_{R_1})P(\langle c_0,0\rangle|r_{R_1})P(\langle c_1,0\rangle|r_{R_1})\dots P(\langle c_n,0\rangle)|r_{R_1})\dots \\
& \dots P(r_{R_{i-1}})P(\langle c_0,0\rangle|r_{R_{i-1}})P(\langle c_1,0\rangle|r_{R_{i-1}})\dots P(\langle c_n,0\rangle)|r_{R_{i-1}}) \\
& \quad P(r_{R_{i+1}})P(\langle c_0,0\rangle|r_{R_{i+1}})P(\langle c_1,0\rangle|r_{R_{i+1}})\dots P(\langle c_n,0\rangle)|r_{R_{i+1}})\dots \\
& \qquad \dots P(r_{R_{jk}})P(\langle c_0,0\rangle|r_{R_{jk}})P(\langle c_1,0\rangle|r_{R_{jk}})\dots P(\langle c_n,0\rangle)|r_{R_{jk}}) \\
& \qquad P(r_t)P(\langle c_0,0\rangle|r_t)P(\langle c_1,0\rangle|r_t)\dots P(\langle c_i,1\rangle|r_t)\dots \\
& \qquad\qquad\qquad\qquad\qquad \dots P(\langle c_n,0\rangle)|r_t)
\end{aligned} \tag{3.22}
$$

これを整理すると,

$$
\begin{aligned}
& P(\langle r_t,c_p,1\rangle,x_j) \\
& \quad = P(n_1)\dots P(n_{i-1})P(n_{i+1})\dots P(n_{jk})P(n_i,v_j,c_p)P(r_{R_1})\dots P(r_{R_{jk}}) \\
& \qquad P(\langle c_0,0\rangle|r_{R_1})\dots P(\langle c_n,0\rangle|r_{R_1})\dots P(\langle c_0,0\rangle|r_{R_{i-1}})\dots P(\langle c_n,0\rangle|r_{R_{i-1}}) \\
& \qquad\quad P(\langle c_0,1\rangle|r_{R_{i+1}})\dots P(\langle c_n\rangle|r_{R_{i+1}})\dots P(\langle c_0,0\rangle|r_{R_{jk}})\dots P(\langle c_n,0\rangle|r_{R_{jk}}) \\
& \qquad\qquad\qquad P(\langle c_0,0\rangle|r_t)\dots P(\langle c_p,1\rangle|r_t)\dots P(\langle c_n,0\rangle|r_t)
\end{aligned} \tag{3.23}
$$

まとめると，

$$P(\langle r_t, c_p, 1\rangle, x_j) = \begin{cases} 3.23, & r_t \text{ な } R \text{ が } x_j \text{ に現れ, その } R \text{ は } i \text{ 番目の名詞と対応している時} \\ 0, & r_t \text{ の値を持つ } R \text{ が } x_j \text{ に現れない場合} \end{cases}$$

さて, ここで 3.23 の確率モデルを構成する個々の確率について考えてみる. $P(n_i)$ の形の確率は名詞の出現する確率である. これは, コーパス上で現れる名詞の頻度を調べることによって最尤推定法により推定が可能である. また, $P(r_t)$ のような R の値の出現確率もコーパスから r_t の頻度を数えることによって推定が可能である. これらの確率はあらかじめ準備することにより既知のものとなる.$P(n_i, v_j, c_p)$ については, 本研究では, これを Rooth[1] の手法による「名詞の動詞に対する係りやすさ」を表す確率として用意する. よってこの確率も既知のものであるとする. 基本モデルの時と同様に,R の値 r_t が現れたときにその値をを持つ名詞がゼロ代名詞の先行詞になる確率 $P(\langle c_p, 1\rangle | r_t)$, およびゼロ代名詞にならない確率が $P(\langle c_p, 0\rangle | r_t)$ が我々にとって未知の確率分布ということになる. よって, $P(\langle c_p, 1\rangle | r_t)$ および $P(\langle c_p, 0\rangle | r_t)$ の分布を, $P(n_i)$, $P(v_j, n_i, c_p)$, $P(r_t)$ の値を用い,EM アルゴリズムの手法によって推定することにする.

3.4 Q 関数の導出

本研究で求めたいのは, R の値 r_t を持つある名詞が格 c_p でゼロ代名詞になる確率分布である. 本研究では, これらを $P(\langle c_p, 1\rangle | r_t)$ と $P(\langle c_p, 0\rangle | r_t)$ による確率分布として求める. これら求めたい確率値をパラメータとし, そのパラメータ集合を θ とする. コーパス上で「観測可能な R の値」の全てのバリエーションが 0 から l までであるとし, 動詞と名詞の格が 0 から n までであるとする. この時パラメータの総数は $2 * (l+1) * (n+1)$ となり, θ は以下のように表される. また, この時,「観測可能な R の値」と格の組合わせのバリエーションを 0 から $m = (l+1) * (n+1) - 1$ とする.

$$\begin{aligned} \theta &= \{P(\langle c_0, 0\rangle | r_0), P(\langle c_0, 0\rangle | r_0), ..., P(\langle c_n, 0\rangle | r_0), P(\langle c_n, 0\rangle | r_0), ... \\ &\quad ..., P(\langle c_0, 0\rangle | r_l), P(\langle c_0, 0\rangle | r_l), ..., P(\langle c_n, 0\rangle | r_l), P(\langle c_n, 0\rangle | r_l)\} \end{aligned} \tag{3.24}$$

このパラメータを含んだコーパス全体の確率は以下の式で与える.

$$\prod_j P_\theta(x_j) \tag{3.25}$$

確率値を対数で表しすとコーパス全体の確率は以下になる.

$$\sum_j log P_\theta(x_j) \tag{3.26}$$

EM アルゴリズムの各ステップで新しく推定するパラメータの集合を $\hat{\theta}$ とすると, 新しく推定されたパラメータを含んだコーパス全体の確率と, 古いパラメータを含んだコーパス全体の確率の差は以下に表される.

$$\sum_j \{logP_{\hat{\theta}}(x_j) - logP_\theta(x_j)\} \quad = \quad \sum_j log\frac{P_{\hat{\theta}}(x_j)}{P_\theta(x_j)} \tag{3.27}$$

今, 考えなくてはならないのは, この新しく更新されたパラメータを含むコーパス全体の確率と, 古いパラメータを含むコーパス全体の確率の差が 0 以上になるようにパラメータの更新方法を導くことである. 新しくなったパラメータによってコーパス全体の生じる確率が下がるという事は, そのパラメータを含む確率モデルには何らかの誤りが含まれているということである. 以下, 一般的な EM アルゴリズムの手法に乗っ取り, 新しいパラメータを含んだコーパス全体の確率と, 古いパラメータを含んだコーパス全体の確率の差を表す Q 関数の導出を行う.

前提として, ある動詞 v_j の前には必ず 1 つのゼロ代名詞が存在していると考えるので, 以下が言える.

$$\sum_{s=0}^{l}\sum_{q=0}^{n} P(\langle r_s, c_q, 1\rangle | x_j) = 1 \tag{3.28}$$

記述の簡明のために, コーパス上で観測可能な R の値 r_s とコーパス上で観測不能なゼロ代名詞と動詞との格 c_q の組 $rc_{sq} = \langle r_s, c_q\rangle$ を用いて記述する. rc の添字 sq は,0 から m まで考える. 0 から m で「観測可能な R の値 r_s と本研究で考慮する格の組み合わせ」の全てのバリエーションを網羅する. これを元に上の式を書き直すと, 以下のようになる.

$$\sum_{sq=0}^{m} P(\langle rc_{sp}, 1\rangle | x_j) = 1 \tag{3.29}$$

3.28,3.29 を用いて 3.27 を変形する.

$$\begin{aligned}
\sum_j log\frac{P_{\hat{\theta}}(x_j)}{P_\theta(x_j)} &= \sum_j \sum_{sq=0}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j) log\frac{P_{\hat{\theta}}(x_j)}{P_\theta(x_j)} \\
&= \sum_j \sum_{sq=0}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j) log\frac{P_{\hat{\theta}}(x_j)}{P_\theta(x_j)} \times 1 \times 1 \\
&= \sum_j \sum_{sq=0}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j) log\frac{P_{\hat{\theta}}(x_j)}{P_\theta(x_j)} \frac{P_{\hat{\theta}}(\langle rc_{sq}, 1\rangle, x_j)}{P_{\hat{\theta}}(\langle rc_{sq}, 1\rangle, x_j)} \frac{P_\theta(\langle rc_{sq}, 1\rangle, x_j)}{P_\theta(\langle rc_{sq}, 1\rangle, x_j)} \\
&= \sum_j \sum_{sq=0}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j) log\frac{P_{\hat{\theta}}(\langle rc_{sq}, 1\rangle, x_j)}{P_\theta(\langle rc_{sq}, 1\rangle, x_j)} \frac{P_\theta(x_j) P_\theta(\langle rc_{sq}, 1\rangle, x_j)}{P_{\hat{\theta}}(\langle rc_{sq}, 1\rangle, x_j) P_\theta(x_j)}
\end{aligned} \tag{3.30}$$

ここでベイズの定理より,

$$\frac{P_\theta(\langle rc_{sq},1\rangle,x_j)}{P_\theta(x_j)}=P_\theta(\langle rc_{sq},1\rangle|x_j)\quad,\quad \frac{P_{\hat\theta}(x_j)}{P_{\hat\theta}(\langle rc_{sq},1\rangle,x_j)}=\frac{1}{P_{\hat\theta}(\langle rc_{sq},1\rangle|x_j)}$$

よって,

$$\begin{aligned}
\sum_j log\frac{P_{\hat\theta}(x_j)}{P_\theta(x_j)} &= \sum_j\sum_{sq=0}^{m}P_\theta(\langle rc_{sq},1\rangle|x_j)log\frac{P_{\hat\theta}(\langle rc_{sq},1\rangle,x_j)}{P_\theta(\langle rc_{sq},1\rangle,x_j)}\frac{P_\theta(\langle rc_{sq},1\rangle|x_j)}{P_{\hat\theta}(\langle rc_{sq},1\rangle|x_j)}\\
&= \sum_j\{\sum_{sq=0}^{m}P_\theta(\langle rc_{sq},1\rangle|x_j)log\frac{P_{\hat\theta}(\langle rc_{sq},1\rangle,x_j)}{P_\theta(\langle rc_{sq},1\rangle,x_j)} \qquad (3.31)\\
&\quad +\sum_{sq=0}^{m}P_\theta(\langle rc_{sq},1\rangle|x_j)log\frac{P_\theta(\langle rc_{sq},1\rangle|x_j)}{P_{\hat\theta}(\langle rc_{sq},1\rangle|x_j)}\} \qquad (3.32)
\end{aligned}$$

ここでジェンセンの不等式を用いる.$P(x)$, $Q(x)$ を確率分布とする.

$$log(x)\le x-1 \qquad (3.33)$$

より次が成り立つ.

$$\begin{aligned}
\sum_x P(x)log\frac{Q(x)}{P(x)} &\le \sum_x\{\frac{Q(x)}{P(x)}-1\}\\
&= \sum_x Q(x)-\sum_x P(x)\\
&= 0 \qquad (3.34)
\end{aligned}$$

ジェンセン　の　不等式

言い換えると,

$$\begin{aligned}
\sum_x P(x)logP(x) &\ge \sum_x P(x)logQ(x)\\
\sum_x P(x)log\frac{P(x)}{Q(x)} &\ge 0 \qquad (3.35)
\end{aligned}$$

よって,3.32 に関しても下のように言える.

$$\sum_{sq=0}^{m}P_\theta(\langle rc_{sq},1\rangle|x_j)log\frac{P_\theta(\langle rc_{sq},1\rangle|x_j)}{P_{\hat\theta}(\langle rc_{sq},1\rangle|x_j)}\ge 0 \qquad (3.36)$$

よって,

$$\sum_j\{logP_{\hat\theta}(x_j)-logP_\theta(x_j)\}\ge\sum_j\sum_{sq=0}^{m}P_\theta(\langle rc_{sq},1\rangle|x_j)log\frac{P_{\hat\theta}(\langle rc_{sq},1\rangle,x_j)}{P_\theta(\langle rc_{sq},1\rangle,x_j)}$$

$$= \sum_j \sum_{sq=0}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j) log P_{\hat{\theta}}(\langle rc_{sq}, 1\rangle, x_j) \\ - \sum_j \sum_{sq=0}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j) log P_\theta(\langle rc_{sq}, 1\rangle, x_j) \quad (3.37)$$

Q 関数を以下のように定義する.

$$Q(\hat{\theta}, \theta) = \sum_j \sum_{sq=0}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j) log P_{\hat{\theta}}(\langle rc_{sq}, 1\rangle, x_j) \quad (3.38)$$

すると, コーパス全体の確率について以下が言える.

$$\sum_j \{log P_{\hat{\theta}}(x_j) - log P_\theta(x_j)\} \geq Q(\hat{\theta}, \theta) - Q(\theta, \theta) \quad (3.39)$$

$Q(\hat{\theta}, \theta) > Q(\theta, \theta)$ となっていれば, 新しいパラメータに更新したとき, コーパス全体の対数尤度は増加することになる. そして, この増加を最大にするためには $Q(\hat{\theta}, \theta)$ を最大にするようなパラメータを求めればいい. 以下, この Q 関数をラグランジェ未定乗数法により $\hat{\theta}$ で偏微分し, 関数 Q を最大にするパラメータの更新式を導く. ここでラグランジェの未定係数法について簡単に説明しておく.

ラグランジェ未定係数法

変数 $x_1, x_2, ..., x_n$ が次の 0 から m までの制約条件を満たすとする. ただし, $m+1 < n$ である.

$$\begin{cases} g_0(x_1, \ldots, x_n) = 0 \\ \quad\quad \vdots \\ g_m(x_1, \ldots, x_n) = 0 \end{cases}$$

この時, これらの制約条件下で関数 $f(x_1, \ldots, x_n)$ を最大化するような $(x_1, \ldots, x_n)$ を求める問題を考える. そのような問題は, 下の関数を $(x_1, \ldots, x_n, \lambda_0, ..., \lambda_m)$ の空間で制約なしで最大化する問題に還元できる.

$$L(x_1, \ldots, x_n, \lambda_0, ..., \lambda_m) \\ = f(x_1, \ldots, x_n) + \sum_{i=0}^{m} \lambda_i g_i(x_1, \ldots, x_n) \quad (3.40)$$

関数 L をラグランジェ関数, $\lambda_0, ..., \lambda_m$ をラグランジェ乗数と呼ぶ. そして関数 L を各変数 x_i で偏微分した式を 0 とおき, 解を求める. その解が定められた制約条件下で関数を最大にする変数である.

$$\begin{cases} \frac{\partial f}{\partial x_1} + \sum_{i=0}^{m} \lambda_i \frac{\partial g_i}{\partial x_1} = 0 \\ \quad\vdots \\ \frac{\partial f}{\partial x_n} + \sum_{i=0}^{m} \lambda_i \frac{\partial g_i}{\partial x_n} = 0 \end{cases}$$

3.5 パラメータの更新式の導出

本研究においては, EM アルゴリズムの各ステップで更新されるパラメータ $\hat{\theta}$ は以下の制約を満たさなくてはならない.

$$\begin{cases} g_0(\hat{\theta}) = P_{\hat{\theta}}(\langle c_0, 0\rangle | r_0) + P_{\hat{\theta}}(\langle c_0, 1\rangle | r_0) - 1 = 0 \\ \quad\vdots \\ g_n(\hat{\theta}) = P_{\hat{\theta}}(\langle c_n, 0\rangle | r_0) + P_{\hat{\theta}}(\langle c_n, 1\rangle | r_0) - 1 = 0 \\ \quad\vdots \\ g_{tp}(\hat{\theta}) = P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t) + P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t) - 1 = 0 \\ \quad\vdots \\ g_{m-n}(\hat{\theta}) = P_{\hat{\theta}}(\langle c_0, 0\rangle | r_l) + P_{\hat{\theta}}(\langle c_0, 1\rangle | r_l) - 1 = 0 \\ \quad\vdots \\ g_m(\hat{\theta}) = P_{\hat{\theta}}(\langle c_0, 0\rangle | r_l) + P_{\hat{\theta}}(\langle c_p, 1\rangle | r_l) - 1 = 0 \end{cases}$$

これら制約の元に, 以下の Q 関数を満たすパラメータをラグランジェの未定係数法により求める.

$$Q(\hat{\theta}, \theta) = \sum_j \{ \sum_{sq=0}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j) log P_{\hat{\theta}}(\langle rc_{sq}, 1\rangle, x_j) \} \tag{3.41}$$

0 から m までの制約に対応するラグランジェ乗数を $\lambda_0 \ldots \lambda_m$ とする. この時ラグランジェ関数 L は以下に表せる.

$$L(\hat{\theta}, \lambda_0, \lambda_1, ..., \lambda_m) = Q(\hat{\theta}, \theta) - \sum_{sq=0}^{m} \lambda_{sq} g_{sq}(\hat{\theta}) \tag{3.42}$$

L を $P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)$ で偏微分する.

$$\begin{aligned} &\frac{\partial L}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} \\ &\quad = \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} \sum_j \{ \sum_{sq=0}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j) log P_{\hat{\theta}}(\langle rc_{sq}, 1\rangle, x_j) \} \\ &\qquad\qquad - \quad \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} \sum_{sq=0}^{m} \lambda_{sq} g_{sq}(\hat{\theta}) \end{aligned} \tag{3.43}$$

ここで, 偏微分の詳細を見るために式を展開していく.

$$
\begin{aligned}
&= \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} \{ \sum_j \{ \sum_{s=0}^{l} \sum_{p=0}^{n} P_\theta(\langle r_s, c_q, 1\rangle | x_j) log P_{\hat{\theta}}(\langle r_s, c_q, 1\rangle, x_j) \\
&\qquad\qquad - \quad \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} \sum_{s=0}^{m} \lambda_{sq} g_{sq}(\hat{\theta}) \\
&= \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} \{ \sum_j \{ P_\theta(\langle r_0, c_0, 1\rangle | x_j) log P_{\hat{\theta}}(\langle r_0, c_0, 1\rangle, x_j) + \ldots \\
&\ldots + P_\theta(\langle r_t, c_p, 1\rangle | x_j) log P_{\hat{\theta}}(\langle r_t, c_p, 1\rangle, x_j) + \ldots \\
&\ldots + P_\theta(\langle r_l, c_n, 1\rangle | x_j) log P_{\hat{\theta}}(\langle r_l, c_n, 1\rangle, x_j) \} \} \\
&\qquad\qquad - \quad \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} \sum_{s=0}^{m} \lambda_{sq} g_{sq}(\hat{\theta})
\end{aligned}
\tag{3.44}
$$

偏微分されるのは新しく更新されるパラメータ $\hat{\theta}$ に関わる部分のみなので,

$$
\begin{aligned}
&= \sum_j \{ P_\theta(\langle r_0, c_0, 1\rangle | x_j) \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} log P_{\hat{\theta}}(\langle r_0, c_0, 1\rangle, x_j) + \ldots \\
&\ldots + P_\theta(\langle r_t, c_p, 1\rangle | x_j) \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} log P_{\hat{\theta}}(\langle r_t, c_p, 1\rangle, x_j) + \ldots \\
&\ldots + P_\theta(\langle r_l, c_n, 1\rangle | x_j) \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} log P_{\hat{\theta}}(\langle r_l, c_n, 1\rangle, x_j) \} \\
&\qquad\qquad - \quad \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} \sum_{s=0}^{m} \lambda_{sq} g_{sq}(\hat{\theta})
\end{aligned}
\tag{3.45}
$$

Q 関数にしろ, 制約条件にしろ, 今は $P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)$ で偏微分しようとしている. すなわち, 微分の対象になる変数は $P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)$ のみであり, その他の変数は係数と同じである. そして,$P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)$ が存在するのは Q 関数の中の $P_{\hat{\theta}}(\langle r_t, c_p, 1\rangle, x_j)$ を含んだ部分と制約条件の $P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)$ と $P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)$ に関する部分だけである. その詳細を追ってやると以下のようになる.

$$
\frac{\partial L}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)}
$$

$$
\begin{aligned}
&= \sum_j \{P_\theta(\langle r_0, c_0, 1\rangle | x_j) \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} log\{P(n_1)\dots P(n_{i-1})P(n_{i+1})\dots P(n_{jk})P(n_i, v_j, c_0) \\
&P(r_0)\dots P(r_{R_{jk}})P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_1})\dots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_1})\dots P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{i-1}})\dots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{i-1}}) \\
&P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{i+1}})\dots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{i+1}})\dots P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{jk}})\dots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{jk}}) \\
&P_{\hat{\theta}}(\langle c_0, 1\rangle | r_0)\dots P_{\hat{\theta}}(\langle c_p, 0\rangle | r_0)\dots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_0)\} + \dots
\end{aligned} \tag{3.46}
$$

$$
\begin{aligned}
&\dots + P_\theta(\langle r_t, c_p, 1\rangle | x_j) \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} log\{P(n_1)\dots P(n_{i-1})P(n_{i+1})\dots P(n_{jk})P(n_i, v_j, c_p) \\
&P(r_{R_1})\dots P(r_{R_{jk}})P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_1})\dots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_1})\dots P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{i-1}})\dots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{i-1}}) \\
&P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{i+1}})\dots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{i+1}})\dots P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{jk}})\dots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{jk}}) \\
&P_{\hat{\theta}}(\langle c_0, 0\rangle | r_t)\dots \underbrace{P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)}_{\text{偏微分の対象となる変数}} \dots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_t)\} + \dots
\end{aligned} \tag{3.47}
$$

$$
\begin{aligned}
&\dots + P_\theta(\langle r_l, c_p, 1\rangle | x_j) \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} log\{P(n_1)\dots P(n_{i-1})P(n_{i+1})\dots P(n_{jk})P(n_i, v_j, c_n) \\
&P(r_{R_1})\dots P(r_{R_{jk}})P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_1})\dots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_1})\dots P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{i-1}})\dots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{i-1}}) \\
&P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{i+1}})\dots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{i+1}})\dots P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{jk}})\dots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{jk}}) \\
&P_{\hat{\theta}}(\langle c_0, 0\rangle | r_l)\dots P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)\dots P_{\hat{\theta}}(\langle c_n, 1\rangle | r_l)\}
\end{aligned} \tag{3.48}
$$

$$
-\lambda_0 \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)}\{P_{\hat{\theta}}(\langle c_0, 0\rangle | r_0) + P_{\hat{\theta}}(\langle c_0, 1\rangle | r_0) - 1\} \tag{3.49}
$$

$$
\vdots
$$

$$
-\lambda_{tp} \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)}\{P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t) + \underbrace{P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)}_{\text{偏微分の対象となる変数}} - 1\} \tag{3.50}
$$

$$
\vdots
$$

$$
-\lambda_m \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)}\{P_{\hat{\theta}}(\langle c_n, 0\rangle | r_m) + P_{\hat{\theta}}(\langle c_n, 1\rangle | r_m) - 1 \tag{3.51}
$$

このうち $P\hat{\theta}(\langle c_p, 1\rangle | r_t)$ を持っているのは,3.47 と 3.50 だけなので, これ以外の部分は偏微分されて 0 になる. よって, この式は結局以下の式と同じことになる.

$$
\begin{aligned}
&\frac{\partial L}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} \\
&\quad = \sum_j \{P_\theta(\langle r_t, c_p, 1\rangle | x_j) \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} log\{P(n_1)\ldots P(n_{i-1})P(n_{i+1})\ldots P(n_{jk})P(n_i, v_j, c_p) \\
&\quad P(r_{R_1})\ldots P(r_{R_{jk}})P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_1})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_1})\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{i-1}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{i-1}}) \\
&\qquad\qquad P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{i+1}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{i+1}})\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{jk}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{jk}}) \\
&\qquad\qquad\qquad P_{\hat{\theta}}(\langle c_0, 0\rangle | r_t)\ldots P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_t)\} \\
&\qquad\qquad -\lambda_{tp} \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} \{P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t) + P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t) - 1\}
\end{aligned}
\tag{3.52}
$$

log の部分を展開すると,

$$
\begin{aligned}
&\frac{\partial L}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} \\
&\quad = \sum_j \{P_\theta(\langle r_t, c_p, 1\rangle | x_j) \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} \{logP(n_1) + \ldots + log(n_{i-1}) \\
&\quad + logP(n_{i+1}) + \ldots + logP(n_{jk}) + logP(n_i, v_j, c_p) + logP(r_{R_1})\ldots logP(r_{R_{jk}})\} + \\
&\quad \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} logP_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_1}) + \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_1}) + \ldots \\
&\quad + \ldots \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} logP_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{i-1}}) + \ldots \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{i-1}}) \\
&\quad + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} logP_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{i+1}}) + \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{i+1}}) + \\
&\quad \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} logP_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{jk}}) + \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{jk}}) \\
&\qquad\qquad + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} logP_{\hat{\theta}}(\langle c_0, 0\rangle | r_t) + \ldots \\
&\qquad\qquad\qquad \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} \underbrace{logP_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)}_{\text{偏微分の対象になる部分}} + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle | r_t)\} \\
&\quad -\lambda_{tp} \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} \{P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t) + \underbrace{P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)}_{\text{偏微分の対象になる部分}} - 1\}
\end{aligned}
\tag{3.53}
$$

$P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)$ 以外の項は $P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)$ で微分して 0 となるので, 結局以下となる.

$$
\frac{\partial L}{\partial P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} = \frac{\sum_j P_\theta(\langle r_t, c_p 1\rangle | x_j)}{P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)} - \lambda_{tp} \tag{3.54}
$$

左辺を 0 とおいて $P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)$ について解くと

$$P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t) = \frac{1}{\lambda_{tp}} \sum_j P_\theta(\langle r_t, c_p, 1\rangle | x_j) \tag{3.55}$$

次に L を $P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)$ で偏微分する. これも L を $P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)$ で偏微分するときと同じように見ていく.

$$\begin{aligned} &\frac{\partial L}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} \\ &\quad = \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} \sum_j \{ \sum_{sq=0}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j) log P_{\hat{\theta}}(\langle rc_{sq}, 1\rangle, x_j) \} \\ &\qquad\qquad - \quad \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} \sum_{sq=0}^{m} \lambda_{sq} g_{sq}(\hat{\theta}) \end{aligned} \tag{3.56}$$

ここで, 偏微分の詳細を見るために式を展開していく.

$$\begin{aligned} &= \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} \{ \sum_j \{ \sum_{s=0}^{l} \sum_{p=0}^{n} P_\theta(\langle r_s, c_p, 1\rangle | x_j) log P_{\hat{\theta}}(\langle r_s, c_p, 1\rangle, x_j) \\ &\qquad\qquad - \quad \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} \sum_{s=0}^{m} \lambda_{sq} g_{sq}(\hat{\theta}) \\ &= \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} \{ \sum_j \{ P_\theta(\langle r_0, c_0, 1\rangle | x_j) log P_{\hat{\theta}}(\langle r_0, c_0, 1\rangle, x_j) + \ldots \\ &\ldots + P_\theta(\langle r_t, c_p, 1\rangle | x_j) log P_{\hat{\theta}}(\langle r_t, c_p, 1\rangle, x_j) + \ldots \\ &\ldots + P_\theta(\langle r_l, c_n, 1\rangle | x_j) log P_{\hat{\theta}}(\langle r_l, c_n, 1\rangle, x_j) \} \} \\ &\qquad\qquad - \quad \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} \sum_{sq=0}^{m} \lambda_{sq} g_{sq}(\hat{\theta}) \end{aligned} \tag{3.57}$$

偏微分されるのは新しく更新されるパラメータ $\hat{\theta}$ に関わる部分のみなので,

$$= \sum_j \{ P_\theta(\langle r_0, c_0, 1\rangle | x_j) \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} log P_{\hat{\theta}}(\langle r_0, c_0, 1\rangle, x_j) + \ldots \tag{3.58}$$

$$\ldots + P_\theta(\langle r_t, c_p, 1\rangle | x_j) \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} log P_{\hat{\theta}}(\langle r_t, c_p, 1\rangle, x_j) + \ldots \tag{3.59}$$

$$\ldots + P_\theta(\langle r_l, c_n, 1\rangle | x_j) \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} log P_{\hat{\theta}}(\langle r_l, c_n, 1\rangle, x_j) \} \tag{3.60}$$

$$- \quad \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} \sum_{sq=0}^{m} \lambda_{sq} g_{sq}(\hat{\theta}) \tag{3.61}$$

今, 微分の対象になる変数は $P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)$ である．Q 関数中で $P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)$ が現れるのは,3.59 の $P_{\hat{\theta}}(\langle r_t, c_p, 1\rangle, x_j)$ 以外の log 項である．というのも,$P_{\hat{\theta}}(\langle r_t, c_p, 1\rangle, x_j)$ 内では必ず $P_{\hat{\theta}}(\langle c_p, 1\rangle|r_t)$ が現れるが, 他の R の値や格についての項では名詞がゼロ代名詞にならないことを表す 0 が現れるからである．すなわち,「観測可能な R の値 r_t」と格 c_p が現れる 3.59 以外の log 項に $P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)$ が出現する．その詳細を追ってやると以下のようになる．

$$
\begin{aligned}
&\frac{\partial L}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)}\\
&\quad= \sum_j \{P_\theta(\langle r_0, c_0, 1\rangle|x_j)\frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} log\{P(n_1)\ldots P(n_{i-1})P(n_{i+1})\ldots P(n_{jk})P(n_i, v_j, c_0)\\
&\quad P(r_{R_1})\ldots P(r_{R_{jk}})P_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_1})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_1})\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle|r_t)\ldots \underbrace{P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)}_{\text{偏微分の対象}}\ldots\\
&\quad \ldots P_{\hat{\theta}}(\langle c_n, 0\rangle|r_t)\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_{i-1}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_{i-1}})\\
&\qquad\qquad P_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_{i+1}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_{i+1}})\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_{jk}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_{jk}})\\
&\qquad\qquad\qquad P_{\hat{\theta}}(\langle c_0, 0\rangle|r_0)\ldots P_{\hat{\theta}}(\langle c_p, 1\rangle|r_0)\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle|r_0)\} + \ldots
\end{aligned}
\tag{3.62}
$$

$$
\begin{aligned}
&\quad \ldots + P_\theta(\langle r_t, c_p, 1\rangle|x_j)\frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} log\{P(n_1)\ldots P(n_{i-1})P(n_{i+1})\ldots P(n_{jk})P(n_i, v_j, c_p)\\
&\quad P(r_{R_1})\ldots P(r_{R_{jk}})P_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_1})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_1})\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_{i-1}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_{i-1}})\\
&\qquad\qquad P_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_{i+1}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_{i+1}})\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_{jk}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_{jk}})\\
&\qquad\qquad\qquad P_{\hat{\theta}}(\langle c_0, 0\rangle|r_t)\ldots P_{\hat{\theta}}(\langle c_p, 1\rangle|r_t)\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle|r_t)\} + \ldots
\end{aligned}
\tag{3.63}
$$

$$
\begin{aligned}
&\quad \ldots + P_\theta(\langle r_l, c_n, 1\rangle|x_j)\frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} log\{P(n_1)\ldots P(n_{i-1})P(n_{i+1})\ldots P(n_{jk})P(n_i, v_j, c_l)\\
&\quad P(r_{R_1})\ldots P(r_{R_{jk}})P_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_1})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_1})\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_{i-1}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_{i-1}})\\
&\qquad\qquad P_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_{i+1}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_{i+1}})\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle|r_t)\ldots \underbrace{P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)}_{\text{偏微分の対象}}\ldots\\
&\quad \ldots P_{\hat{\theta}}(\langle c_n, 0\rangle|r_t)\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_{jk}})P_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_{jk}})\ldots\\
&\qquad\qquad\qquad P_{\hat{\theta}}(\langle c_0, 0\rangle|r_{r_l})\ldots P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)\ldots P_{\hat{\theta}}(\langle c_n, 1\rangle|r_l)\}
\end{aligned}
\tag{3.64}
$$

$$
-\lambda_0 \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)}\{P_{\hat{\theta}}(\langle c_0, 0\rangle|r_0) + P_{\hat{\theta}}(\langle c_0, 1\rangle|r_0) - 1\} \tag{3.65}
$$

$$
\vdots
$$

$$
-\lambda_{tp} \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)}\{\underbrace{P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)}_{\text{偏微分の対象}} + P_{\hat{\theta}}(\langle c_p, 1\rangle|r_t) - 1\} \tag{3.66}
$$

$$\vdots$$

$$-\lambda_m \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} \{P_{\hat{\theta}}(\langle c_n, 0\rangle | r_m) + P_{\hat{\theta}}(\langle c_n, 1\rangle | r_m) - 1 \tag{3.67}$$

このうち $P\hat{\theta}(\langle c_p, 0\rangle | r_t)$ を持っていないの部分は偏微分されて 0 になる．Q 関数に関しては 3.63 以外は残ることになる．よってこの式は結局以下の式と同じことになる．

$$\begin{aligned}
&\frac{\partial L}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} \\
&\quad = \sum_j \{P_\theta(\langle r_0, c_0, 1\rangle | x_j) \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} log\{P(n_1)\ldots P(n_{i-1})P(n_{i+1})\ldots P(n_{jk})P(n_i, v_j, c_0) \\
&\quad P(r_{R_1})\ldots P(r_{R_{jk}})P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_1})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_1})\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle | r_t)\ldots P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)\ldots \\
&\quad \ldots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_t)\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{i-1}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{i-1}}) \\
&\qquad\quad P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{i+1}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{i+1}})\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{jk}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{jk}}) \\
&\qquad\qquad P_{\hat{\theta}}(\langle c_0, 1\rangle | r_0)\ldots P_{\hat{\theta}}(\langle c_p, 0\rangle | r_0)\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_0)\} + \ldots
\end{aligned} \tag{3.68}$$

$$\begin{aligned}
&\quad \ldots + P_\theta(\langle r_l, c_n, 1\rangle | x_j) \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} log\{P(n_1)\ldots P(n_{i-1})P(n_{i+1})\ldots P(n_{jk})P(n_i, v_j, c_l) \\
&\quad P(r_{R_1})\ldots P(r_{R_{jk}})P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_1})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_1})\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{i-1}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{i-1}}) \\
&\qquad\quad P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{i+1}})\ldots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{i+1}})\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle | r_t)\ldots P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)\ldots \\
&\quad \ldots P_{\hat{\theta}}(\langle c_n, 0\rangle | r_t)\ldots P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_{jk}})P_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_{jk}})\ldots \\
&\qquad\qquad P_{\hat{\theta}}(\langle c_0, 0\rangle | r_{r_l})\ldots P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)\ldots P_{\hat{\theta}}(\langle c_n, 1\rangle | r_l)\}
\end{aligned} \tag{3.69}$$

$$-\lambda_{tp} \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} \{P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t) + P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t) - 1 \tag{3.70}$$

log 項を展開すると，

$$\begin{aligned}
&\frac{\partial L}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} \\
&\quad = \sum_j \{P_\theta(\langle r_0, c_0, 1\rangle | x_j) \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} \{logP(n_1) + \ldots + logP(n_{i-1}) + logP(n_{i+1}) + \ldots \\
&\quad \ldots + logP(n_{jk}) + logP(n_i, v_j, c_0) + logP(r_{R_1})\ldots logP(r_{R_{jk}})\} + \\
&\quad \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} logP_{\hat{\theta}}(\langle c_0, 0\rangle | r_{R_1}) + \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle | r_{R_1}) + \ldots \\
&\qquad \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} logP_{\hat{\theta}}(\langle c_0, 0\rangle | r_t) + \ldots \\
&\qquad \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} \underbrace{logP_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)}_{\text{偏微分の対象}} + \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle | r_t) + \ldots
\end{aligned}$$

$$
\begin{aligned}
&\ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_{i-1}}) + \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_{i-1}}) \\
&+ \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_{i+1}}) + \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_{i+1}}) + \\
&\ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_{jk}}) + \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_{jk}}) \\
&+ \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_0)} logP_{\hat{\theta}}(\langle c_0, 0\rangle|r_0) + \ldots \\
&\ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_0)} logP_{\hat{\theta}}(\langle c_p, 0\rangle|r_0) + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle|r_0) + \ldots \\
&\ldots + P_{\theta}(\langle r_l, c_n, 1\rangle|x_j) \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} \{logP(n_1) + \ldots + logP(n_{i-1}) + logP(n_{i+1}) + \ldots \\
&\ldots + logP(n_{jk}) + logP(n_i, v_j, c_l) + logP(r_{R_1}) \ldots logP(r_{R_{jk}})\} + \\
&\frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_1}) + \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_1}) + \ldots \\
&\ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_{i-1}}) + \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_{i-1}}) \\
&+ \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_{i+1}}) + \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_{i+1}}) + \\
&\quad \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_0, 0\rangle|r_t) + \ldots \\
&\qquad \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} \underbrace{logP_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)}_{\text{偏微分の対象}} + \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle|r_t) + \ldots \\
&\ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_0, 0\rangle|r_{R_{jk}}) + \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle|r_{R_{jk}})\} \\
&+ \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_0)} logP_{\hat{\theta}}(\langle c_0, 0\rangle|r_l) + \ldots \\
&\quad \ldots + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_0)} logP_{\hat{\theta}}(\langle c_p, 0\rangle|r_l) + \frac{\partial}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)} logP_{\hat{\theta}}(\langle c_n, 0\rangle|r_l)\} \\
&\qquad -\lambda_{tp} \frac{\partial}{\partial P_{\hat{\theta}}(c_p, 0\rangle|r_t)} \{\underbrace{P_{\hat{\theta}}(\langle c_p, 0\rangle|r_t)}_{\text{偏微分の対象}} + P_{\hat{\theta}}(\langle c_p, 1\rangle|r_t) - 1\}
\end{aligned}
\tag{3.71}
$$

最終的には, $P(\langle c_p, 0\rangle|r_t)$ 以外の項は $P(\langle c_p, 0\rangle|r_t)$ で微分して 0 となる. $P_{\theta}(\langle r_t, c_p, 1\rangle|x_j) logP_{\hat{\theta}}(\langle r_t, c_p, 1\rangle, x_j)$ を持つ項以外は微分が行われるので, 以下のようになる.

$$\frac{\partial L}{\partial P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} = \sum_j \{ \frac{\sum_{sq=0}^{tp-1} P_\theta(\langle rc_{sq}, 1\rangle | x_j)}{P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} + \frac{\sum_{sq=tp+1}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j)}{P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)} \} - \lambda_{tp} \tag{3.72}$$

左辺を 0 とおいて $P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)$ について解く.

$$P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t) = \frac{1}{\lambda_{tp}} \sum_j \{ \sum_{sq=0}^{tp-1} P_\theta(\langle rc_{sq}, 1\rangle | x_j) + \sum_{sq=t+1}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j) \} \tag{3.73}$$

$P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t)$ と $P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t)$ は足し合わせて 1 になるという制約を満たすので,

$$\frac{1}{\lambda_{tp}} \sum_j \{ P_\theta(\langle r_t, c_p, 1\rangle | x_j) + \sum_{sq=0}^{tp-1} P_\theta(\langle rc_{sq}, 1\rangle | x_j) + \sum_{sq=tp+1}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j) \} - 1 = 0 \tag{3.74}$$

整理すると,

$$\frac{1}{\lambda_{tp}} \sum_j \{ \sum_{sq=0}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j) \} = 1 \tag{3.75}$$

動詞の前に出現した名詞には必ずゼロ代名詞があるという前提なので, 以下が言える.

$$\sum_{sq=0}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j) = 1 \tag{3.76}$$

よって,3.75 について下が言える.

$$\frac{1}{\lambda_{tp}} \sum_j \{1\} = 1 \tag{3.77}$$

動詞は V 個あるので

$$\frac{V}{\lambda_{tp}} = 1 \tag{3.78}$$

よって,

$$\lambda_{tp} = V \tag{3.79}$$

よって, パラメータの更新式は以下のように表される.

$$P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t) = \frac{1}{V}\sum_j P_\theta(\langle rc_{tp}, 1\rangle | x_j) \tag{3.80}$$

$$P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t) = \frac{1}{V}\sum_j \{\sum_{sq=0}^{tp-1} P_\theta(\langle rc_{sq}, 1\rangle | x_j) + \sum_{sq=tp+1}^{m} P_\theta(\langle rc_{sq}, 1\rangle | x_j)\} \tag{3.81}$$

ここで

$$\begin{aligned} P_\theta(\langle rc_{tp}, 1\rangle | x_j) &= \frac{P_\theta(\langle rc_{tp}, 1\rangle, x_j)}{P(x_j)} \\ &= \frac{P_\theta(\langle rc_{tp}, 1\rangle, x_j)}{\sum_{sq=0}^{m} P(\langle rc_{sq}, 1\rangle, x_j)} \end{aligned} \tag{3.82}$$

よって, 基本モデルのパラメータの更新式は最終的に以下のようになる.

$$P_{\hat{\theta}}(\langle c_p, 1\rangle | r_t) = \frac{1}{V}\sum_j \frac{P_\theta(\langle rc_{tp}, 1\rangle, x_j)}{\sum_{sq=0}^{m} P(\langle rc_{sq}, 1\rangle, x_j)} \tag{3.83}$$

$$P_{\hat{\theta}}(\langle c_p, 0\rangle | r_t) = \frac{1}{V}\sum_j \{\frac{\sum_{sq=0}^{ts-1} P_\theta(\langle rc_{sq}, 1\rangle, x_j)}{\sum_{sq=0}^{m} P(\langle rc_{sq}, 1\rangle, x_j)} + \frac{\sum_{sq=ts+1}^{m} P_\theta(\langle rc_s, 1\rangle, x_j)}{\sum_{sq=0}^{m} P(\langle rc_{sq}, 1\rangle, x_j)}\} \tag{3.84}$$

ただし,

$$\begin{aligned} &P(\langle rc_{sq}, 1\rangle, x_j) = P(\langle r_t, c_p, 1\rangle, x_j) \\ &\quad = P(n_1)\ldots P(n_{i-1})P(n_{i+1})\ldots P(n_{jk})P(n_i, v_j, c_p)P(r_{R_1})\ldots P(r_{R_{jk}}) \\ &\qquad P(\langle c_1, 0\rangle | r_{R_1})\ldots P(\langle c_n, 0\rangle | r_{R_1})\ldots P(\langle c_1, 0\rangle | r_{R_{i-1}})\ldots P(\langle c_n, 0\rangle | r_{R_{i-1}}) \\ &\qquad\quad P(\langle c_1\rangle | r_{R_{i+1}})\ldots P(\langle c_n\rangle | r_{R_{i+1}})\ldots P(\langle c_1, 0\rangle | r_{R_{jk}})\ldots P(\langle c_n, 0\rangle | r_{R_{jk}}) \\ &\qquad\qquad P(\langle c_1, 0\rangle | r_t)\ldots P(\langle c_p, 1\rangle | r_t)\ldots P(\langle c_n, 0\rangle | r_t) \end{aligned} \tag{3.85}$$

第4章　実験

本章では, 本研究で提案した手法の有効性を調べるために行った実験について述べる. 実験を行うための手順は以下のようになる.

1 新聞記事から R の値を抽出する

2 抽出した R の値と Rooth の手法による「名詞の動詞に対する係りやすさを」表す確率分布を用いて, EM アルゴリズムに基づいて導出した本研究のパラメータ更新式を元に R に関する確率分布の学習を行う

3 ゼロ代名詞の照応関係の正解を与えたコーパスを作成する

4 得られた R に関する確率分布を用いてゼロ代名詞の照応解析を行う

以下, これらの手順ごとに説明を与える.

4.1　R の抽出

これは学習データの作成に相当する作業である. 本研究では, 無作為に選ばれた読売新聞 3210 記事に対して学習データの作成を行った. この作業の手順は以下のような流れにしたがって行われる.

1 新聞記事に記事の区切り, 先頭の段落の区切りを与え, それを構文解析する

2 構文解析結果に対して, 文節ごとの名詞句と動詞句を ID に汎化する

3 1 と 2 の結果から記事中の全ての動詞に対して, その前出現している名詞ごとに R の値を作成する

1 では, まず集められた新聞記事に対して, 記事の区切りを与える. 本研究が対象とするゼロ代名詞の範囲は, 文内照応を含める前方照応であり, 外界照応は扱わない. R に関する確率分布もその範囲で学習を行うので, 学習データも記事を単位とする. また同時に, 先頭の段落の区切りを与える. そして, JUMAN で形態素解析を行う. その際,JUMAN コマンドに対して-i オプションを与え, 記事の境界と先頭の段落の境界を示す印[1]を無視して形態

[1]'articlebound' および 'endofhead' という文字列で与えた

素解析を行った[2]. その後に既存の構文解析器 [10] で構文解析している. この係り受け解析の結果は後で行う R の値の抽出にも用いられる. また, ここで行っている構文解析の目的は, 記事中に存在する文を文節ごとに区切る作業であり, 区切られた文節が名詞句なのか動詞句なのかを区別する作業である.

2 は区切られた文節を汎化する作業である. 本研究では, ゼロ代名詞の照応解析および R に関する確率分布の学習に Rooth の手法による,「名詞の動詞に対する係りやすさ」を表す確率分布を使用している. しかし, 実際には, 本研究では, Rooth の手法を日本語に対して拡張した鳥澤の確率分布 [3] を用いている. 鳥澤のシステムでは文節に現れる名詞句や動詞句は全て ID で区別される. 例えば, コーパス中に現れる全ての「食べる」は「25768」で区別する. ここでは, 鳥澤が「名詞の動詞に対する係りやすさ」を表す確率分布を学習した際に作成した ID に従って, 構文解析結果の文節を置き換えていく. 本研究での学習の際にも,「名詞の動詞に対する係りやすさ」を表す確率分布を使用するからである. また, この時, 語尾の助詞によって動詞句は区別される. 例えば,「食べる：られる」と「食べる」は区別される.「られる」などの助詞が動詞についた場合, 動詞のとる格フレームは大きく変化するので, この処理は重要である. 例えば, 以下のように変る.

「食べる」のガ格の格要素は「人間」や「動物」に属する名詞がなりやすい

「食べる：られる」のガ格の格要素は「食べ物」に属する名詞がなりやすい

また, 連続名詞は, 連続名詞を構成する後ろの形態素の意味が, 連続名詞の意味の中心になるように切り出される. これは例えば, 次のように汎化される.「渡辺:泰三:政府:代表」なら「*:政府:代表」に汎化される. この「*」であるが, 正規表現のメタ文字と似た働きをするものである. この場合は,「政府:代表」も「ベラルーシ:政府:代表」も「*:政府:代表」を表す ID で置き換わることになる.

3 の R の値の作成は, 学習コーパス上で観測可能な素性のみを行う. 具体的には以下のように行う.

- 名詞と動詞の間の文間距離 d： 文の区切りごとに, それまで記事にいくつ文節があるかをカウントし, その数字を元に, それぞれの文節が記事の中の何文目に現れるかを計算する. そして, 記事に現れた全ての動詞に対して, その動詞より前の文節で出現した名詞との間に, 文がいくつあるかを計算する

- 名詞に後置する助詞 $c1$： 構文解析の出力結果には, 文節を構成する自立語と付属語は分けて出力されている. ここから, 名詞に後置する助詞が何であるかを抽出し, 助詞ごとの ID に汎化する

- 名詞が連体節に含まれているかどうか $rentai$： 名詞が連体修飾節に含まれているかどうかの判断を, 本研究では,「名詞が係っている動詞が名詞に係っているならその

[2]-i オプションが有効なのは一種類の文字列だけなので, 結局は JUMAN 出力を自前のスクリプトで修正した

名詞は連体修飾節に含まれている」と判断し, 構文解析の出力結果から, 名詞が連体修飾節に含まれるかそうでないかを抽出している.

- 名詞が先頭の段落に含まれているか否か $head$: 記事の中で先頭の段落の印が現れるまでにいくつの文節があるかをカウントし, それを元に記事中の名詞に対して先頭の段落に含まれているか, そうでないかを抽出する
- 動詞の前に現れる R ごとの ID $eachid$: 以上の素性を抽出した後に, 記事中の動詞ごとに, 同じ R の値を持っている場合には動詞に近い方から番号を与えていく

本研究では, これらの作業によって, 読売新聞 3210 記事中の全ての動詞に対して, その動詞の前に出現する全ての名詞との位置関係を抽出する. 記事中に動詞は 97681 回出現し,R の値は 8289552 個作成された. 具体的な R の値としては, 下のような数字列が作成される.

> 1:3:01:2 →名詞と動詞の間は 1 文ある. 名詞の助詞「は」である. 名詞は連体節に含まれない. 先頭の段落には含まれる. 動詞の前に現れる 2 番目の「1:3:01」である.

4.2 学習

２章で導出した R に関する確率分布の推定法を実装し, $P(\langle c_p, 1\rangle | r_t)$ の分布を推定した. 4.1 に読売新聞 3210 記事で推定した時の, $P(\langle c_p, 1\rangle | r_t)$ を確率値の高いものから上位 15 を示す.

確率値が上位のものは, 名詞と動詞との文間距離が 0 のものが多い. 全体の傾向として距離が遠くなるほど確率の値が下がっており, 名詞と動詞の間の距離が近いほどゼロ代名詞が起りやすいという現象をつかめているのではないかと思われる. また,「30 日東京に行く」の「30 日」は無格として動詞に係っていると解釈されるが, このような助詞のない名詞が, 無格やニ格で動詞に係る場合は, 距離による確率値の低下は他の場合も小さくなっている. これは, 日付などの名詞は距離を超えてゼロ代名詞として現れやすいということを表していると思われる.

次に, 名詞に後置する助詞とゼロ代名詞の格との関係であるが, ゼロ代名詞の格と同じ格助詞を持つ名詞が先行詞として参照されやすいということが示されている. また, 注目すべきは, 提題助詞「は」は「ガ格」で係りやすいということを確率分布として捉えられているということである. また, 日付などの助詞のない名詞は, ニ格で現れやすいという現象[3]も捉えられている.

[3]「30 日東京に行く」は「30 日に東京に行く」とも言える

順位	文間距離	名詞の助詞 → 格 (c_p)	連体節	先頭の段落	番号	$P(\langle c_p, 1\rangle \mid r_t)$
1	0	に → に	含まれない	含まれない	1	0.08795
2	0	を → を	含まれない	含まれない	1	0.08733
3	0	が → が	含まれない	含まれない	1	0.05024
4	0	を → を	含まれる	含まれない	1	0.04198
5	0	に → に	含まれる	含まれない	1	0.03365
6	0	助詞無し → 無格	含まれない	含まれない	1	0.03133
7	0	を → を	含まれない	含まれる	1	0.02511
8	0	は → が	含まれない	含まれない	1	0.02428
9	1	助詞無し → 無格	含まれない	含まれない	1	0.02274
10	0	に → に	含まれない	含まれない	1	0.02197
11	0	の → に	含まれない	含まれない	1	0.01995
12	1	の → に	含まれない	含まれる	1	0.01917
13	1	に → に	含まれない	含まれない	1	0.01750
14	2	助詞無し → 無格	含まれない	含まれない	1	0.01573
15	0	を → を	含まれる	含まれる	1	0.01449

表 4.1: 名詞と動詞の位置関係に関する確率値の上位 15

順位	文間距離	名詞の助詞 → 格 (c_p)	連体節	先頭の段落	番号	$P(\langle c_p, 1\rangle \mid r_t)$
40	1	助詞なし → に	含まれない	含まれない	1	0.0050

また,「の → に」の確率値が上位に入っている.「の」は「管理署のもみの木」のように他の名詞を修飾するために用いられる. 学習データ内での出現頻度も高く, 確率値が大きくなったものと思われる.

また, 全体の傾向として, 先頭の段落に含まれる場合の確率値は, 先頭の段落に含まれない場合の確率値よりも低くなっている. 本研究では, 先頭の段落に含まれる名詞は, 記事の話題の中心になっており, ゼロ代名詞になりやすいという予測を立てていたが, 確率分布としてはその傾向を捉えることはできなかった.

4.3 正解コーパスの作成

4.3.1 ゼロ代名詞のタグをつけたコーパス

本研究では, 照応解析実験に先だって, 実験で正解として利用するゼロ代名詞タグ付コーパスを作成した. 本研究では, 学習データを作成した読売新聞 3210 記事とは別の記事から無作為に抽出した読売新聞 10 記事分に対してゼロ代名詞についてのタグを与え, ゼロ代名詞タグ付けコーパスとした. 本研究で用いたゼロ代名詞タグ付コーパスとは実質的に

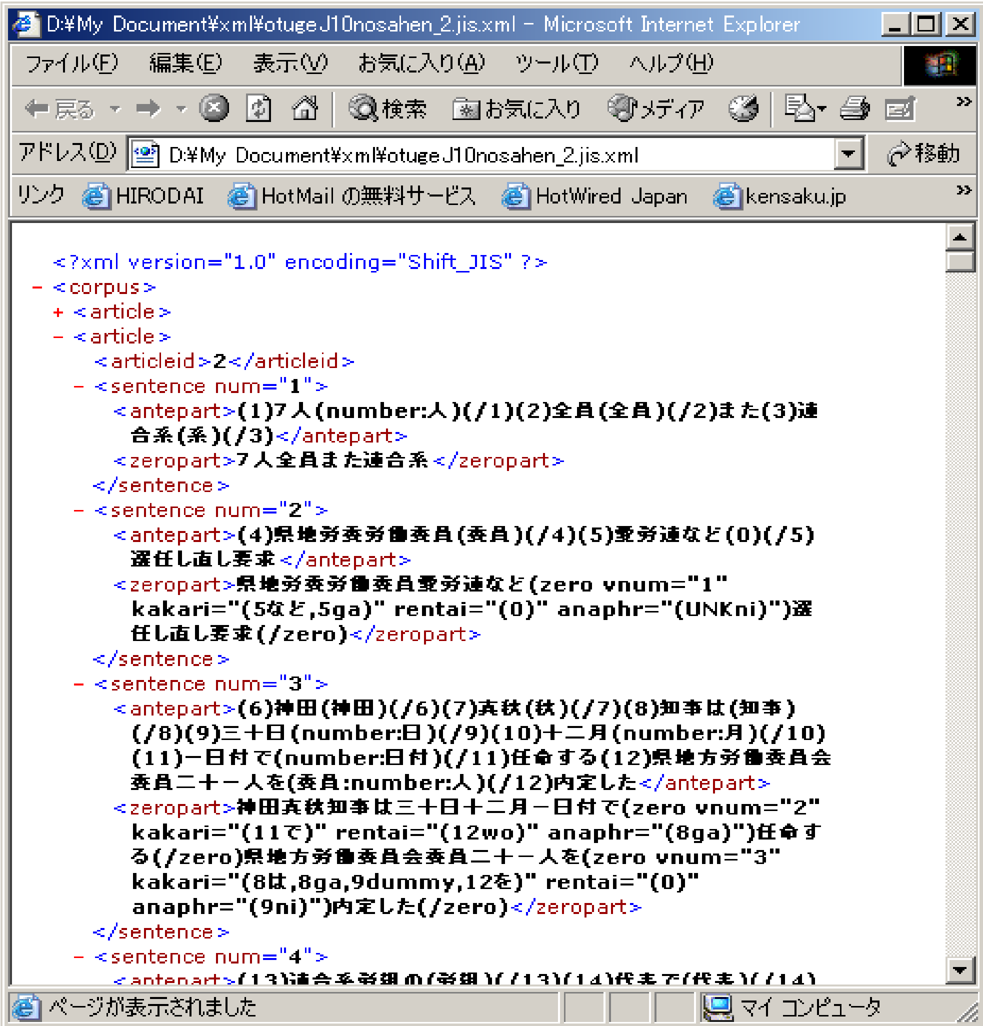

図 4.1: IE6.0 で表示したゼロ代名詞タグ付コーパス

XML ファイルである．タグ付けとは実質的には, 記事中の動詞句を囲む XML タグの属性値を埋めていくという作業になる. 記事中の全ての名詞句文節には, 記事内における一意な番号が与えられ, ゼロ代名詞の先行詞候補となる名詞句は, この番号によって区別される. そして, 動詞に対して, 何番の名詞句が何格でゼロ代名詞になっているかをタグの属性値として書き込む.

4.1 は IE6.0 で表示したタグをつけた後のコーパスである．ゼロ代名詞のタグ付けは以下に注意して行った.

先行詞の単位とゼロ代名詞の種類

本研究では, ゼロ代名詞の先行詞は文節内の自立語部分とした. 本研究では, ゼロ代名詞とは目には見えなくなってしまった格要素であるとする. よって, 消えてしまったゼロ代名詞は文節の形, すなわち自立語と付属語の形で存在すると考える. ただし, 前方照応ではないが, 文脈上明らかにゼロ代名詞があるという場合には, 外界照応を示す記号を動詞のタグ内の属性に付加した. 例えば,2001 年 9 月 11 日以降の記事において「テロ事件が起った」という表現があったならば,「起った」の「デ格」には「ニューヨーク」が現れていると思われる. この場合に対してタグ付を行うとすると, このタグの属性は「デ格」で外界照応によってゼロ代名詞が現れている記号が付加される.

格の交替について

同じ文中の同じ名詞は, 一度しか格要素にならないということに注意してタグを付けた. 例えばこれは下のような例に当てはまる.

誤ったゼロ代名詞のタグ付け：労働組合で (労働組合が) 作る連合会議

この場合「作る」という動詞には, 格助詞「が」で係っている格要素は存在しない.「作る」のガ格は「労働組合」である. しかし, この時, ガ格のゼロ代名詞の先行詞が「労働組合」とすることは間違いである. 同じ文中に「労働組合」はすでに出現しており, 本研究では「労働組合」は「作る」のゼロ代名詞にはならないと考える. この場合は格助詞「で」によって「作る」のガ格が作られている. このような現象を格の交替という [14].

組織やその代弁者などがゼロ代名詞になる場合

例えば, 下のような例に当てはまる.

例:島根銀行の頭取は「不景気に強い体質を作り上げたいと」語った.

この場合の「作り上げたい」のガ格は「島根銀行」と「頭取」のどちらの意味にもとれる. このような場合には, 両方が正解となるようにタグを付けた.

名詞の同一性について

ゼロ代名詞の先行詞が指し示すものが, 記事中で複数の表現で現れている場合, その名詞が前方照応の先行詞候補になるならば, 本研究ではその全ての多様な表現を正解としてタグ付を行った. 「名詞が複数の表現で現れる」とは, 以下のような例である.

- 固有名詞が記事後半で短く表現される：渡辺泰三政府代表→渡辺代表

- 代名詞が用いられる：島根銀行→同行,30 日→同日

例えば,「島根銀行」が記事の 1 文目に現れ,「同行」が 3 文目に現われたとする. この「島根銀行」と「同行」は同じものを指し示している. この時, 4 文目にゼロ代名詞が発生しており, システムが「島根銀行」をゼロ代名詞の先行詞として判断したとすると, 本研究ではこれは正しいと判断する.

ニ格について

動詞が持てる格要素の数の上限はガ格, ヲ格, ニ格ともにそれぞれ一つが原則である. ただし, ニ格に関しては 2 つ以上の格要素を持つことを許す.

- 例：政府代表は (30 日に)(東京に) 到着する.

この場合の「30 日に」と「東京に」は同じニ格であるが, 意味としては「時間」と「目的地」となり異なる. また, 格の交替が起った場合も「30 日に」は「30 日」と省略することができるが,「東京に」の「に」は省略不能であり, むしろ「東京へ」のように「へ格」で格の交替が起る. このようにニ格は意味が広く, 意味の異なるニ格でゼロ代名詞が生じる場合は, 本研究では別々の格要素として扱う.

4.3.2 ゼロ代名詞タグ付け結果

本研究で作成したゼロ代名詞タグ付けコーパスについて報告する. 下の図はタグ付きコーパス中の動詞に現れる前方照応のガ格, ヲ格, ニ格のゼロ代名詞の数である. ニ格のゼロ代名詞に関しては,「(30 日に)(東京に) 行った」のようなニ格の意味が異なるゼロ代名詞が生じた場合に 2 つの格要素を許す. この場合には「行った」には 2 つのニ格のゼロ代名詞が生じていると判断する. また, この時（）の中は未知語を含む場合の数である. 以下の結果からゼロ代名詞のほとんどがガ格とニ格で占められているということが分かる.

ガ格	ヲ格	ニ格	計
64(66)	2(2)	64(66)	112(115)

ゼロ代名詞コーパス内のゼロ代名詞数

また, ニ格のゼロ代名詞で格要素が 1 つだけでゼロ代名詞になっているものと, 格要素が 2 つでゼロ代名詞になっている内訳は以下のようになっている. （）の中は未知語を含む場合の数である.

ニ格のゼロ代名詞の内訳

格要素が 1 つの場合	格要素が 2 つの場合	ニ格合計
30(31)	16(16)	46(47)

	ガ格	ヲ格	ニ格	計	5 記事に占める割合
距離 0	6(7)	0(0)	5(5)	11(12)	0.1692(0.1714)
距離 1	5(5)	0(0)	11(12)	16(17)	0.2462(0.2429)
距離 2	4(4)	0(0)	2(2)	6(6)	0.0923(0.0857)
距離 3	5(6)	0(0)	3(3)	8(9)	0.1231(0.1285)
距離 4	4(4)	0(0)	1(1)	5(5)	0.0769(0.0714)
距離 5	1(1)	0(0)	1(1)	2(2)	0.0307(0.0286)
距離 6	1(1)	0(0)	2(2)	3(3)	0.0462(0.0429)
距離 7	4(4)	0(0)	2(2)	6(6)	0.0923(0.0857)
距離 8	4(5)	0(0)	2(2)	6(7)	0.0923(0.1000)
距離 9	1(1)	0(0)	0(0)	1(1)	0.0154(0.0143)
距離 10	1(2)	0(0)	0(0)	1(2)	0.0154(0.0286)
距離 11 以上	0(0)	0(0)	0(0)	0(0)	0.0(0.0)

表 4.2: 文間距離ごとの正解先行詞数

以上はゼロ代名詞の発生回数を示したが, 次はゼロ代名詞の先行詞になることのできる記事中の正解数を示す. 本研究では, 組織や代弁者がゼロ代名詞になる場合や, 同一の対象を指し示す表現が記事中に複数現れる場合には, そのゼロ代名詞の先行詞に複数の名詞を正解としており, 正解の数はゼロ代名詞の発生数よりも大きくなる.（）の中は未知語を含む場合の数である. また, 表の 3 段目は,1 つのゼロ代名詞の出現に対する正解先行詞数の平均である.

	ガ格	ヲ格	ニ格	計
正解数	134(140)	6(6)	81(82)	221(228)
正解先行詞数の平均	2.094(2.121)	3.0(3.0)	1.267(1.2424)	1.9732(1.9828)

ゼロ代名詞コーパス内の正解先行詞数

これより,1 つのゼロ代名詞の出現に対する正解先行詞数はおよそ 2 であるということが分かる.

4.3.3 ゼロ代名詞タグ付コーパスの正解の詳細

ゼロ代名詞タグをつけたコーパスの正解について分析を行う. しかし, 本研究で作成したコーパスは 10 記事分であり, これを行ってゼロ代名詞照応解析実験を行った. このコーパス全てを分析すると実験の公正さを失う恐れがある. コーパスの正解に合わせて実験内容を変えたという謗を受けることになる. 本研究では,10 記事のうち 5 記事について, コー

パスの詳細を分析し, その分析結果で得た知見を 10 記事による実験に生かすという方法をとった. 以下が 5 記事における, ゼロ代名詞数と先行詞正解数である.

	ガ格	ヲ格	ニ格	計
ゼロ代名詞	20(21)	0(0)	22(23)	42(44)
正解数	36(40)	0(0)	29(30)	65(70)
正解先行詞数の平均	1.8000(1.9048)	0(0)	1.3181(1.3043)	1.5476(1.5909)

5 記事内のゼロ代名詞と正解先行詞数

また,4.2 に文間距離ごとの正解の数を表示する. 表の全体に占める割合とは, 各距離ごとの正解の数を全ての正解の数で割った数字である. ここで注目すべきは, 距離 3 までの正解が全体の 63% を占めるということである.

次に先頭の段落に含まれている正解数と, 先頭の段落に含まれていない正解数を示す.

	ガ格	ヲ格	ニ格	計	5 記事に占める割合
先頭の段落に含まれる	21(23)	0(0)	25(26)	46(49)	0.7077(0.7000)
先頭の段落に含まれない	15(17)	0(0)	4(4)	19(21)	0.2923(0.3000)

5 記事内で先頭の段落に含まれる正解数と含まれない正解数

正解先行詞の 70% が先頭の段落に含まれており, 多くのゼロ代名詞が先頭の段落を参照していることが分かる. 河原ら [8] は, 省略の指示対象候補の 98.6% が以下の範囲にあると報告している.

- 対象用言がある文中
- 前文の重要な要素
- 記事の最初の文の重要な要素

河原らのゼロ代名詞の範囲は, 文内の後方照応もカバーしており, 前方照応だけを扱う本研究とは異なるが, 本研究では河原らの指示対象候補の範囲を以下のように言い直して調査を行った.

- 文間距離 0
- 文間距離 1
- 先頭の段落に含まれる

すると, 正解数は以下のようになった.

正解についている助詞	ガ格	ヲ格	ニ格	計	5 記事に占める割合
に	0(0)	0(0)	5(5)	5(5)	0.0769(0.0714)
から	0(0)	0(0)	1(1)	1(1)	0.0154(0.0143)
が	3(3)	0(0)	3(3)	6(6)	0.0923(0.0857)
の	5(6)	0(0)	1(1)	6(7)	0.0923(0.1000)
や	4(4)	0(0)	2(2)	6(6)	0.0923(0.0857)
は	20(23)	0(0)	4(4)	24(27)	0.3692(0.3857)
を	2(2)	0(0)	1(1)	3(3)	0.0462(0.0429)
助詞なし	2(2)	0(0)	12(13)	14(15)	0.2154(0.2143)

表 4.3: 正解名詞に後置助詞とゼロ代名詞の関係

ガ格	ヲ格	ニ格	計	5 記事に占める割合
30(33)	0(0)	27(28)	57(61)	0.8769(0.8714)

5 記事の内の河原らによる指示対象の範囲の限定

98.6% とはいかないまでも,87.7% の正解をカバーしていることが分かる. ここで文間距離を 3 以下のものを含めると以下のようになる.

ガ格	ヲ格	ニ格	計	全体に占める割合
33(37)	0(0)	28(29)	61(66)	0.9384(0.9429)

5 記事のうちの本研究による指示対象の範囲の限定

この指示対象の範囲の限定によりおよそ 94% の先行詞の正解を限定できたことになる.

最後に 4.3 に,5 記事中の正解先行詞を後置する助詞ごとに分類してみた.「は」のついた先行詞の多くがガ格で現れている. また,「助詞なし」で現れる先行詞はニ格で現れやすい. これは日付などの名詞が当てはまるものと思われる.

4.4 照応解析実験

読売新聞 10 記事で作成したゼロ代名詞タグ付コーパスに対して, 3210 記事で学習して得た「R に関する確率分布」と Rooth の手法で得られた「名詞の動詞に対する係りやすさ」を表す確率分布を用いてゼロ代名詞照応解析実験を行った. 実験は関ら [4] [5] の行った「ゼロ代名詞の先行詞特定の実験」に準じる. すなわち, 以下のような実験を行う.

- ゼロ代名詞の先行詞特定の実験：ゼロ代名詞の位置と格が正しく検出されたという前提で, システムにゼロ代名詞の先行詞を出力させて正解かどうかを判定する. 本研究で得た確率分布 $P(\langle c_p, 1\rangle | r_t)$ の性能を検証する実験である.

4.4.1 先行詞を決定するスコアの算出法

システムは, 格 c のゼロ代名詞が出現している動詞があったとき, その動詞の前に出現している先行詞候補に対して「格 c での先行詞になりやすさ」を表すスコアを計算し, そのスコアの高いものほど先行詞になりやすいと判断する. 動詞 v があり, そこに格 c のゼロ代名詞があった時の名詞 n_i スコアは, 名詞 n_i の R の値 r_t を用いて以下のように算出される. このスコアによるゼロ代名詞照応解析が本研究手法である.

$$n_i\text{のスコア} = P(\langle c, 1\rangle | r_t) \times P(n_i, v, c) \tag{4.1}$$

$P(\langle c, 1\rangle | r_t$ は本研究で推定した「位置関係の値 r_t を持つ名詞が格 c のゼロ代名詞の先行詞になる」確率である. ただし, r_t の素性のうち, $eachid$ は全て 1 として扱った. これは, 最も信頼のできる $\langle d, c1, rentai, head\rangle$ による確率値を用いるためである. $P(n_i, v, c)$ は Rooth の手法による「名詞の動詞に対する係りやすさ」すなわち「名詞 n_i が v に格 c で係っている」ことを表す確率である. これを以後「係り受け確率」と呼ぶ. また, 本研究では $P(n_i, v, c)$ のみでスコアを計算する精度と, $P(\langle c, 1\rangle | r_t)$ のみでスコアを計算する精度をそれぞれ取得し, $P(\langle c, 1\rangle | r_t) \times P(n_i, v, c)$ でスコアを計算する精度と比較する.

4.4.2 精度の算出方法

精度は正解率として以下の式で算出する.

$$\text{正解率} = \frac{\text{システムの出力が正解した回数}}{\text{コーパスに現れたゼロ代名詞の数}} \tag{4.2}$$

- 出力した先行詞候補が正解する:格 c のゼロ代名詞があり, その格 c のゼロ代名詞の格要素が k 個 だった時に, システム先行詞候補の出力の上位 k 位までに正解が含まれている場合

関ら [5] は, システムの出力上位 1 位に正解が含まれている場合をカウントして正解率が 50.7% になったと報告している. しかし, 本研究では, ニ格の格要素が複数存在することを認めており, かつ R に関する確率分布をゼロ代名詞のタグのついていない素のコーパスから取得している. そして, その複数で現れる格要素を区別する手段を持たない. 現状では, 複数の格要素ニ格のゼロ代名詞先行詞候補として一種類の順位付けしかできないのである. すなわち, 関らの精度の出し方だと, 「(30 日に)(東京に) 行った」のようにゼロ代名詞が現れている場合, この時のシステムの出力が「1 位:30 日」「2 位:東京」となったとしても, 2 位の「東京」は正解にはカウントできないのである. よってこの場合は, ニ格の格要素は 2 種類あるということから, システムの上位 2 位までに正解に含まれているとして, 「1 位:30 日」「2 位:東京」の両方を正解としてカウントする.

4.4.3 意味クラスの粒度

本研究では, ゼロ代名詞照応解析に Rooth[1] の手法による確率分布を用いている. 実際には, 本研究では Rooth の手法を日本語に対して拡張した鳥澤の手法 [3] による確率分布を用いている. 鳥澤の確率分布では 500 個の意味クラスによる確率分布と 2500 個の意味クラスによる確率分布の 2 種類が存在する.500 クラスに比べて 2500 クラスは, より細かく単語を分類することができる. 例えば,「温泉」という単語で $P(温泉, c, v)$ の値が最大になるような c と v の組を探してみたところ, 500 クラスならば,「温泉がある」となり,2500 クラスならば「温泉に入る」となった. また,「マグロ」だと 500 クラスで「マグロを飲む」となり, 2500 クラスならば「マグロ (助詞なし) 採る」となる. クラス 500 では「食べ物も飲み物もひとまとめ」な比較的粒度の粗いクラスであり, クラス 2500 はそれよりも粒が細かいクラスである.

4.4.4 実験結果

以下は照応解析実験の結果である. 正規スコアとは, $P(\langle c, 1\rangle | r_t) \times P(n_i, v, c)$ でスコアを計算し, 係り受け確率のみとは $P(n_i, v, c)$ のみでスコアを計算し, 位置関係のみとは $P(\langle c, 1\rangle | r_t$ のみでスコアを計算した場合である.

	ガ格	ヲ格	ニ格	計	正解率
正規スコア	26(25)	0(0)	16(16)	42(40)	0.375(0.3565)
係り受け確率のみ	12(11)	0(0)	16(17)	28(28)	0.25(0.2435)
位置関係のみ	22(25)	0(0)	5(5)	27(30)	0.24107(0.26089)

ゼロ代名詞の同定

河原ら [8] は先行詞候補を選ぶ際に, その先行詞候補の位置によって選別したと報告している. これと同様に本研究でも, 先行詞を位置によって選別することを考える. 本研究では以下のように行う.

- 文間距離が 3 以下の候補か, 先頭の段落に含まれている先行詞以外は候補から除く

さらに, 本研究では正解コーパスの半分を調査した際に得た知見から以下の制約を課す. ここでは以下のような処理を行った.

- 「の」がついている先行詞候補は除く
- 係り受け確率が 1.0e-08 より小さいものは 1.0e-08 で置き換える

「の」は名詞を接続して次の名詞を修飾する働きがある. 本来ゼロ代名詞として現れるのは, 被修飾名詞である. しかし, 文脈によっては「の」が後置する名詞もゼロ代名詞にな

ることができる. 例えば「300 人の各国代表」という表現があった場合,「300 人」と「各国代表」は同じ対象を指し示しているとも解釈可能であるが, ゼロ代名詞として参照されるのは「各国代表」であると考える. 読み手にとっては, 「300 人」は「の」で「各国代表」に接続することにより「各国代表」の量を表していると解釈されるからである. しかし,「営林署のもみの木」という表現だと,「営林署」と「もみの木」は別の対象を指し示している. この時, 「営林署」を参照するようなゼロ代名詞があり,「営林署」の表現が「営林署のもみの木」しかない場合には, ゼロ代名詞の先行詞は「営林署」しかありえない. このように「の」が後置する名詞がゼロ代名詞の先行詞になる場合は文脈に左右されやすい. しかし,$P(\langle c_p, 1\rangle | r_t)$ の分布の調査の際に, 名詞に「の」が後置する R の確率がかなり高くなっていた. これにより, 先行詞決定の時に他の正しい正解が選ばれるのが妨げられているのではないかと考え, 本研究では「彼の」の「彼」のように「の」が続く名詞を先行詞の候補から外した.

また, 係り受け確率が極端に低い場合には,1.0e-08 という値で係り受け確率を置き換えた. これは, 係り受け確率が極端に低い場合にも $P(\langle c_p, 1\rangle | r_t)$ の値によって正解が選ばれる場合があると考えたからである. 結果は以下のようになった.

	ガ格	ヲ格	ニ格	計	正解率
正規スコア	27(29)	0(0)	19(19)	46(48)	0.41071(0.41391)
係り受け確率のみ	15(14)	0(0)	16(17)	32(28)	0.28571(0.269565)
位置関係のみ	26(28)	0(0)	8(6)	34(34)	0.30357(0.29565)

制約を導入した場合のゼロ代名詞同定結果

この 41.07% がゼロ代名詞同定実験で本研究で出した最高の精度であり, 関らの 50.7% に大きく劣る結果となった.

最後に係り受け確率を 2500 クラスのもので計算した結果を以下に示す. これは上記の制約で先行詞候補を除いていない. また, スコアの算出に R に関する確率のみを用いる精度は変らない.

	ガ格	ヲ格	ニ格	計	正解率
正規スコア	22(21)	0(0)	16(16)	38(37)	0.339285(0.321739)
係り受け確率のみ	13(12)	0(1)	18(18)	32(31)	0.28571(0.269565)
位置関係のみ	24(25)	0(0)	5(5)	29(30)	0.258929(0.2608696)

係り受け確率をクラス 2500 で行ったゼロ代名詞の同定

ここで注目すべきなのは, 係り受け確率のクラスを 500 から 2500 に増やした時に, 係り受け確率のみで精度を算出する場合には精度が上がるが, 正規スコアは精度が下がるという点である. ゼロ代名詞照応解析において, R の確率は粒の粗いクラスの係り受け確率と相性がよいと言える. 目の粗い係り受け確率で先行詞候補にだいたいの「あたり」をつけた上で, R によって先行詞候補を選んだほうが, 粒の細かい係り受け確率を用いる場合よ

りも精度は良くなるということが分かる. また, 係り受け精度のみによるゼロ代名詞同定の最高精度は,2500 クラスの係り受け確率を用い, 先行詞候補から係り受け確率が極端に低いものを 1.0e-08 で置き換えた場合であり, ゼロ代名詞数 112 に対して正解数は 33 となり, 精度は 0.29642 であった. この精度は, 関らの実験における意味モデルだけの精度とほぼ同等であった.

第 5 章　考察と今後の課題

5.1　考察

本研究で得た結果と河原ら [8], 関ら [4][5] による関連研究を比較する.

5.1.1　関らの手法との比較

本研究は手法では, 確率の値でゼロ代名詞の照応解析を行おうとしているわけであるから, 関らの研究に近い. 本研究と関らの研究で使用する確率モデルの形も似ている. 関らの統語モデルは本研究での「名詞と動詞の位置関係に関わる確率分布」であり, 意味モデルは Rooth の手法による確率分布に相当する. ただ, 関らの統語モデルで扱われている素性は「先行詞とゼロ代名詞の文間距離」と「先行詞が連体修飾節に含まれているかどうか」と「先行詞の助詞とゼロ代名詞の格」であり, これらはそれぞれ独立の確率分布として推定され, ゼロ代名詞の照応解析に用いられている. 本研究ではこれに加えてさらに「先行詞が先頭の段落に含まれているか」という素性があり[1], 確率変数には素性の組からなる一つのベクトルを与えた. しかし, 精度の上では大きな差が現れた. 本研究での最高精度の 41.07% に対して関らの最高精度は 50.7% と本研究は大きく引き離されている. 本研究における, Rooth の確率分布のみを用いた場合の精度は 29.6% であり, 関らの意味モデルのみの精度は 29.5% と, 本研究と関らの精度はほぼ同等なのである. しかし, 本研究の位置関係のみの最高精度の 30.4% に対して, 関らの統語モデルのみの精度はは 45.8% と本研究の精度を大幅に上回っている. この差によって, 本研究と関らの研究の最高精度の差が作られていると考えられる. 本研究の位置関係の確率と関らの統語モデルの最も大きな違いは, 本研究の確率分布はタグ無しコーパスから R の値ごとに, Rooth の手法による確率分布や名詞と動詞の位置関係の出現確率の値などを用いて推定したのに対して, 関らの確率はゼロ代名詞のタグのついたコーパスから, ゼロ代名詞の起った場合に関わる頻度をカウントすることによって推定している. 本研究では, ゼロ代名詞とは関係のない確率分布からゼロ代名詞照応解析のための確率分布を推定しているのである. それゆえに, 現時点では本研究で推定した確率分布はゼロ代名詞の現象を捉えきれていないと考えられる. 係り受けや R の値が生じる現象とゼロ代名詞が現れる現象の裏には, 本研究では扱うことができなかった隠れ素性が存在すると考えられる. 本研究で作成したゼロ代名詞タグ付コーパスを調査したところ, ゼロ代名詞の約 70% が先頭の段落に正解を持っているにも関わらず,4.1

[1] eachid は実験では全て 1 として扱ったのでここでは割愛する.

に示したように, 確率分布ではその性質を捉えきれていない. 今後タグ付けコーパスの量を増やして, ゼロ代名詞の生じる場合というものをさらに検討しなくてはならない.

5.1.2 河原らの手法との比較

次に本研究と河原の研究を比較する. 河原らのシステムの用言に対する適合率は 68.7% であり, これはゼロ代名詞の検出も含めての精度である. 河原らの論文 [8] によると, ゼロ代名詞検出の際に行った格フレームの選択に関する記述について「ほとんどの用言の用法に一致したものであり, 格フレームが有効に用いられた」とあることから, ほとんどのゼロ代名詞の検出に成功していると見ることができる. ゼロ代名詞が正しく検出できた状況での精度としての 68.7% は, 関らの精度よりも 18.0%, 本研究よりも 27.6% 高いことになる. 確率の値を用いてゼロ代名詞の照応解析を行う手法を大きく引き離していることになる. この高い精度を産み出した要因として, 先行詞が存在している範囲を制約によって制限し, 98.6% の先行詞が存在する範囲内から先行詞候補を選ぶことに成功している点がまずあげられる. 本研究でも文間距離や先頭の段落の区別を付けることで先行詞候補を絞ることを行っているが, 河原らはもっと細かく 1 文内における「重要な要素」を定義し, 主節の用言に係る格要素, スコープの長い従属節内の格要素, 提題助詞「は」のついている名詞句から先行詞候補を選んでいる. また, 格フレーム辞書を用いることで, すでに係っている格要素とゼロ代名詞の格要素の両方を用いてスコア付することを行っており, 先行詞と動詞の 1 対 1 の対応しか捉えきれなかった確率値を用いる本研究や関らの研究を引き離しているものと考えられる.

5.1.3 本研究の現時点での結論

本研究は他の関連研究に比べて精度が低く, 今後に渡って改善を続けていかなくてはならない. しかし, 本研究は関らや河原の研究とは異なり, 人手によるシソーラスを一切使用していない. Rooth の手法による確率分布も EM アルゴリズムによってタグ無しコーパスから学習されたものである. そして, Rooth による確率分布のみを用いてゼロ代名詞照応解析を行った場合の精度が 29.6% であり, Rooth の手法による確率分布と名詞と動詞の位置関係に関わる確率分布の両方を用いたときの精度が 41.07% であるので, タグ無しのコーパスからの学習だけによって約 11% の精度の向上に成功したことになる. これにより, 教師無し学習結果を用いてさらに教師無し学習することにより, ゼロ代名詞照応解析の精度が向上する見通しが得られたことになる. 以上から, 本研究では, 本研究の手法によって取得された確率分布により, ゼロ代名詞照応解析の精度の向上が確かめられたと結論づける.

5.2 今後の課題

最後に, 本研究の今後の課題を述べる.

- ゼロ代名詞タグ付コーパスを増やす:　本研究では, 読売新聞 10 記事文にゼロ代名詞タグを付与し, その内 5 記事から知見を得つつ, ゼロ代名詞の同定の実験を行った. しかし,10 記事は必ずしも十分な量ではない. 河原らは毎日新聞 20 記事, 関らは毎日新聞 30 記事に対してゼロ代名詞のタグ付を行っている. 今後, 本研究でもタグ付コーパスを増やし, その上で最初にタグをつけた 10 記事全てから, 本研究で得た確率分布のゼロ代名詞照応解析におよぼす作用について詳細を検討する予定である.

- R の新しい素性を加えての確率分布の再学習:　本研究で推定した確率分布は, 精度の上からもゼロ代名詞の特性を完全に捉えたものであるとは言えない. よって,R の値の素性を新たに加え再学習を試みる必要がある. 具体的には以下のような素性が考えられる.

  - 河原は先行詞候補の絞り込みを文間だけでなく, 文内においても行っている. これを参考に, 本研究でも用言に直接係っている格要素とそうでない名詞を区別し, 再学習を行うことが考えられる.
  - 関らの手法の確率モデルでは, 統語に関する素性ごとに独立の確率分布が推定され, ゼロ代名詞の照応解析に用いられている. これに対して, 本研究では, 名詞と動詞の位置関係は 1 つのベクトルの値として与えられている. 本研究の精度は関らの精度に劣っており, 本研究でも素性を分離して再学習を行い, 関らのモデルと本研究のモデルの差を検証する必要がある.
  - R の値の素性にクラスを導入することも考えられる. これは名詞の種類によってゼロ代名詞の現れ方に違いがでることを捉える試みである. 例えば, 記事の先頭付近で現れた「日付」や「組織」などに属する名詞は, 記事を通して出現することが多い. これらの素性を新たに R の値に加え, 再学習を行う.

- 複数の格要素を考慮したゼロ代名詞照応解析モデル:　河原らは大規模コーパスから自動的に構築された格フレーム辞書を用いて複数の格要素を同時に考慮するゼロ代名詞照応システムを開発し, 高い精度をあげている. 鳥澤は Rooth の確率モデルを拡張し,「名詞 n_1 が格 c_1 で, 名詞 n_2 が格 c_2 でそれぞれ動詞 v に係る」確率分布を定義している [2]. これを本研究にも導入し, 本研究でも複数の格要素を同時に考慮するモデルの構築を目指す.

謝辞

主テーマ指導教官の鳥澤健太郎助教授には, 研究の全般において多大なご意見, ご助言を頂きました. 心より御礼申し上げます. 東条敏教授には多くのご意見を頂きました. 深く感謝いたします. 永田祐一助手には多くのご助言を頂きました. 深く感謝いたします. 知識工学講座の皆さんには, 公私に渡りとてもお世話になりました. ありがとうございます.

参考文献

[1] Mats Rooth, Stefan Riezler, Detlef Prescher, Glenn Carroll, and Franz Beil. Inducing a semantically annotated lexicon via em-based clustering. In *Proceedings of 37th Annual Meeting of the ACL*, 1999

[2] Kentaro Torisawa. A Nearly Unsurpervised Learning Method for Automatic Paraphrasing of Japanese Noun Phrases. In *Proceedings of the Workshop on Automatice Paraphrasing*, pp.63-72, Tokyo, Japan, 2001

[3] Kentaro Torisawa. An unsupervised method for canonicalization of Japanese postpositons. In *Proceedings of 6th Natural Lauguage Processing Pacific Rim Symposium (NLPRS2001)*, pp.211-218, Tokyo, Japan, 2001

[4] Kazuhiro Seki, Atsushi Fujii, Tetsuya Ishikawa. A Probabilistic Model for Japanese Zero Pronoun Resolution Integrating Syntactic and Semantic Features. In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium (NLPRS2001)*, pp.403-410, 2001

[5] Kazuhiro Seki, Atsushi Fujii, Tetsuya Ishikawa. A Probabilistic Mothod for Analyzing Japanese Anaphora Integrating Zero Pronoun Detection and Resolution. In *Proceedings of the 19th International Conference on Conputational Linguistics (COLING 2002)*, pp.911-917, 2002

[6] 森田真樹, 長尾真. 用例や表層表現を用いた日本語文章中の指示詞, 代名詞, ゼロ代名詞の指示対象の推定. 自然言語処理, Vol.4, No.1, pp.87-109,1997

[7] 河原大輔, 黒橋禎夫. 用言と直前の格要素を組を単位とする格フレームの自動構築. 自然言語処理 Vol.9, No.1, pp.3-19,2002

[8] 河原大輔, 黒橋禎夫. 自動構築された格フレーム辞書に基づく省略解析. 言語処理学会 第7回年次大会, 2001

[9] 中岩浩巳. コーパスからの省略補完ルール獲得環境. 情報処理学会研究報告, NL-138, pp.31-38, 2000

[10] Hiroshi Kaneyama, Kentaro Torisawa, Yutaka Mitsuishi, and Jun'ichi Tsujii. A hybrid japanese parser with hand-crafted grammar and statistics. In *Proceedings of COLING 2000, pp.411-417,2000*

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum lilelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society, 39(B):1-38, 1977

[12] Malilyn Walker, Masayo Iida, Sharon Cote. Japlanese Discourse and the Process of Centering. Computational Linguistices, Vol.20, No.2, pp.193-232, 1994

[13] 竹井みつこ, 高田美佳, 相沢輝昭. 日本語ゼロ代名詞補完のためのグローバルトピックの役割. 情報処理学会研究報告, NL-135, pp.71-78, 2000

[14] 益岡隆志, 仁田義男, 郡司隆男, 金水敏. 文法. 岩波講座 言語の科学 5, 岩波書店, 1996