

日本語書き言葉を対象とした人間の自然な省略検出の分析

飯田龍¹ 橋本力¹ 鳥澤健太郎¹ 黒橋禎夫² 乾健太郎³ 宮尾祐介⁴ 柴田知秀² 笹野遼平⁵

¹ 情報通信研究機構ユニバーサルコミュニケーション研究所

² 京都大学大学院情報学研究科 ³ 東北大学大学院情報科学研究科

⁴ 国立情報学研究所 ⁵ 東京工業大学精密工学研究所

1 はじめに

代名詞等の指し先(先行詞)を特定する照応解析、主語等の省略を補完する省略解析は自然言語処理の形態素・構文解析のような表層的な解析処理と質問応答、自動要約、機械翻訳等の応用処理をつなぐ必須の処理であり、人工知能の分野で古くから研究が行われている[4, 6]。このうち、指し元の表現(照応詞)が代名詞や名詞句である照応解析、もしくは、照応詞と先行詞の指す実体の同一性までを問題とする共参照解析では、照応詞と先行詞の文字列の一致情報を利用したり、二つの表現の意味的な互換性を WordNet[3] のような言語資源を利用することで、比較的容易に関係を同定可能であることが知られている。一方、省略解析の場合は照応詞側が省略され、文字列一致や意味的な互換性等の情報を利用できないため、(代)名詞間の照応解析、もしくは共参照解析と比較して、自動解析の性能が低いことが知られており、この問題に着目したさらなる研究の成果が渴望されている。また、日本語では、主語等の省略が頻繁に起こるため、省略解析を高精度で解析できることにより、その後に想定されている応用処理の性能が向上することが期待できる。このような背景より、日本語を対象とした省略解析の研究が盛んに進められてきた[5, 10, 7, 2, 1]。

これらの省略解析の既存研究の多くは、省略された項(ゼロ代名詞)とその先行詞の関係(以降、省略関係)が人手によりアノテーションされたタグ付きコーパスを利用しており、コーパス中に出現する省略関係の特徴を学習することで自動解析モデルを構築している。このため、自動解析の性能はコーパス内で定義された省略関係の定義に強く依存することになる。例えば、述語項関係がガ格やヲ格といった表層格関係でアノテーションされている京都大学テキストコーパス[8]やNAISTテキストコーパス[11]では、項が係り受け関係にない場合には、同一文内もしくは前方の文からその項を補完するというアノテーションが行われており、このアノテーションの結果が省略解析の問題として扱われている。ただし、省略補完に対して制約がまったく無いと、網羅的にアノテーションされた省略関係の中には、先行詞として補完できる表現がゼロ代名詞から離れた位置に偶然出現している場合等も含まれることになり、この結果、省略解析の問題集合に解析が極端に困難な事例が含まれることになる。また、述語の項をアノテーションする際に探索範囲に制限をかけないと、先行詞として補完すべき項の検出に揺れが生じるという問題もある。さらに、京都大学テキストコーパスとNAISTテキストコーパスでは、前者が述語の交替まで含めた表層格に対して項をアノテ

ションしているのに対し、後者は述語の原形に対して表層格をアノテーションするという異なるアノテーションの仕様を採用しており、どのように省略解析の問題を設計するかについては、研究者間でコンセンサスが得られていない状況にあるといえる。

このような背景に対し、例えば、工藤ら[9]が高性能な機械翻訳の実現を目的として、文章中に記述されていない一人称や二人称等の代名詞を省略補完するための訓練事例を作成する等、各タスクに合致した省略解析の研究が進められており、これらの取り組みは個々のタスクの性能向上のために重要な課題となる。一方で、それらの個別の取り組みとは独立に、人間の言語理解を明らかにするために、人間が何を省略とみなすかという観点から省略解析の問題を考えることも可能である。そこで、本研究では「人間が自然に省略と判断する省略の現象が省略解析の問題としてコンセンサスの得られる問題集合となる」という仮説に基づき、人間が共通に自然に省略だと判断できる箇所が存在するのか、また、その省略にはどのような特徴があるのかを調べることで、省略関係の問題として採用する基準について検討する。具体的には、複数人のアノテーション作業者に最低限の省略関係の概要を説明した上で、同一の文章集合にアノテーションさせ、その結果を分析することで、人間が自然に省略だと判断が一致した現象にどのようなものがあるかを調査する。まず、2節でアノテーション作業の概要を説明し、3節でアノテーションの結果について示す。さらに、4節でアノテーションされた結果の特徴を人手で分析し、また、NAISTテキストコーパスとの違いについて調査した結果について報告する。最後に5節でまとめと今後の課題を議論する。

2 自然な省略検出課題

省略箇所を人間が検出したデータを収集するために、自然言語処理に関係するアノテーションに未習熟な作業員8名に対し、NAISTテキストコーパスから抽出した100記事(5,634述語)¹を対象に、以下に示す省略検出作業を行わせ、省略箇所を検出させた。

省略検出のインストラクション: 省略を自然に認識してほしい一方、今回作業を行う作業員は自然言語アノテーションの素人であるため、こちらが予測しないような明らかにおかしなアノテーションを行う可能性がある。そこで、作業の前に必須格と任意格の違いの概要の説明や、単文中に格要素が全て埋まっている場合や連体修飾の関係にある表現が格要素として埋まる場合には省略関係と

¹この100記事中にアノテーションされた省略の個数は、ガ格が1,122、ヲ格が139、二格が47、合計1,308であった。

| | B | C | D | E |
|---|---------|----|----|----|
| 1 | 述語/文脈 | ガ格 | ヲ格 | ニ格 |
| 2 | 太郎は公園に | | | |
| 3 | 行った | x | x | x |
| 4 | 。 | | | |
| 5 | EOS | | | |
| 6 | そこで、花子と | | | |
| 7 | 会った | 太郎 | x | x |
| 8 | 。 | | | |
| 9 | EOS | | | |

図 1: アノテーション作業画面

みなさないことや、逆に「太郎は公園に行った。そこで、花子と会った。」の2文目で動詞「会う」の主語が省略されている例を示し、このような場合には省略関係を認定することを伝えた²。また、交替をともなう述語については、述語原形に対して無理に省略関係を認定されるのを抑制するために、京都大学テキストコーパスと同様に、受け身、使役に関する格交替を含む表現に対して省略関係を認定する作業とした。さらに、複文、つまり、一つの文に述語が複数存在する場合、既存のコーパスで採用されているような「格要素が係り受け関係にない場合には省略関係を検出する」と同じようなアノテーション結果を得られるのかを調査するために、あえてどのように省略関係を認定すべきかを指定せずに作業を行わせた。アノテーションでは、NAISTテキストコーパス作成時の予備調査で頻出したガ格、ヲ格、ニ格の省略関係のみをアノテーションの対象とした。

Excelを用いたアノテーション作業: アノテーションの作業そのものに不慣れな作業員に対して習熟が必要となる専門的なアノテーションツールを導入すると、そのツールの操作性のために誤った作業結果が生じる可能性がある。そこで、本研究では、図1に示すように、一般に広く利用されているExcelを用い、Excel上で文章を述語の前後で改行した形式で示し、その述語に対してガ格、ヲ格、ニ格の省略を検出させ、その検出した省略の先行詞を文章内から抜粋するという作業を行わせた³。

3 アノテーション結果

作業員8名が同一の100記事を対象にアノテーションを行った結果、省略を検出した個数を表1に示す。この結果を見てわかるように、作業員ごとに省略検出の個数は大きく異なることがわかる。特に、作業員A1が他の作業員よりも多く省略を検出している。これは後述する表4でも示すように、A1がアノテーション作業員の中で最も多く文内の省略を付けており、その結果、全アノテーション数が多くなってしまったと考えられる。

また、省略の先行詞を補完する際、多くのアノテーションでは、文内に出現する複数の述語のいずれか一箇所に項を補完しているのに対し、場合によっては二箇所以上記入されている場合も存在した。そこで、作業員ごとにそのような二箇所以上の補完が起こった頻度を調査した結果を表2に示す。この表より、作業員A1、A3、A6が全体の平均よりも多くアノテーションしており、その結

²単文、連体修飾関係についても同様に具体例を用いて説明した。
³つまり、本アノテーション作業では、先行詞が文章内に存在する文脈照応(endophora)のみがアノテーションの対象となる。

表 1: 人間による省略検出数

| アノテータ | 全ての格 | ガ格 | ヲ格 | ニ格 |
|-------|------|-----|----|-----|
| A1 | 1061 | 689 | 95 | 277 |
| A2 | 356 | 268 | 29 | 59 |
| A3 | 379 | 355 | 4 | 20 |
| A4 | 598 | 461 | 45 | 92 |
| A5 | 234 | 195 | 27 | 12 |
| A6 | 651 | 508 | 27 | 116 |
| A7 | 471 | 253 | 60 | 158 |
| A8 | 507 | 384 | 52 | 71 |
| 平均 | 532 | 389 | 42 | 101 |

表 2: 同一文内で複数の同一の先行詞を補完した個数

| アノテータ | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | 平均 |
|-------|----|----|----|----|----|----|----|----|------|
| 個数 | 56 | 21 | 46 | 11 | 11 | 60 | 19 | 13 | 29.6 |

果、全体のアノテーションの個数が多くなったと考えられる。

さらに、同一文章集合に対するNAISTテキストコーパス1.5版にアノテーションされた省略の個数⁴は、全ての格で1,395、ガ格が1,210、ヲ格が137、ニ格が48となっており、ヲ格の省略がニ格の省略よりも多いのに対し、一方で、本作業の平均ではニ格の省略がヲ格の省略よりも多くなっていることがわかる。NAISTテキストコーパスではアノテーションされていないが、本アノテーション作業で少なくとも1名がニ格をアノテーションした事例を調査したところ、例えば、「通告する」、「移住する」、「出馬する」等、ニ格が必須格として埋まることが望ましいと考えられる述語にアノテーションされていることがわかった。ニ格の必須、任意の判定は非常に困難であるため、一人の作業員ではそのアノテーションに漏れや揺れが起こることが知られているが、このアノテーション結果は、その検出漏れや作業の揺れを複数人でアノテーションすることで統制できる可能性を示唆している。

次に、8名のアノテーション作業員が作業した結果の省略検出の一致数⁵、つまり、作業員横断的に同一の省略箇所を検出した個数を一致数とともに計上した結果を表3に示す。この結果より、8名全員がすべて省略だと判定した箇所はほとんど存在しないことがわかる。また、少なくとも一人以上がセルに記入して省略検出をした個数が1,978であるのに対し、例えば、作業員数の半数以上である5名以上で一致した個数は231であり、全体の12%しか一致していないことがわかる。

次に、ガ格、ヲ格、ニ格の箇所に記入された先行詞の文字列を手がかりに、省略とその先行詞がどのような位置関係にあるかを調査した。具体的には、記入された先行詞の文字列が文章中のどの位置に出現するかで関係を分類した。この際、省略と同一文内に出現(intra)省略より前の文に出現(pre)省略より後の文に出現(post)文字列が合致しない⁶(other)という順序で決定的に出現位置を決定した。この手順に従い、各アノテーション作業員の作業結果全体、もしくは格ごとに出現位置の頻度を調査した結果を表4に示す。この結果のうち、特に

⁴省略の個数を求める際、本アノテーション作業で対象としている文脈照応のみを計上した。

⁵ここでは述語の項に対して先行詞が補完されたか否かのみに基づいて一致数を求めた。このため、補完された先行詞が同一の表現であるかは考慮していない点に注意されたい。

⁶これは特定の作業員が作業内容を誤解し、先行詞として補完すべき表現を要約して短い句として生成、もしくは文章内の内容を解釈し、そこに出現していない表現を補完したために起こった。

表 4: 省略関係の種類とアノテーションの個数

| | 全事例 | | | | ワ格 | | | | ヲ格 | | | | 二格 | | | |
|----|-------|-----|------|-------|-------|-----|------|-------|-------|-----|------|-------|-------|-----|------|-------|
| | intra | pre | post | other | intra | pre | post | other | intra | pre | post | other | intra | pre | post | other |
| A1 | 373 | 674 | 9 | 5 | 239 | 441 | 7 | 2 | 33 | 58 | 2 | 2 | 101 | 175 | 0 | 1 |
| A2 | 50 | 285 | 1 | 20 | 33 | 217 | 1 | 17 | 5 | 24 | 0 | 0 | 12 | 44 | 0 | 3 |
| A3 | 10 | 368 | 1 | 0 | 8 | 346 | 1 | 0 | 0 | 4 | 0 | 0 | 2 | 18 | 0 | 0 |
| A4 | 73 | 507 | 13 | 5 | 58 | 387 | 11 | 5 | 4 | 39 | 2 | 0 | 11 | 81 | 0 | 0 |
| A5 | 42 | 190 | 2 | 0 | 35 | 159 | 1 | 0 | 4 | 22 | 1 | 0 | 3 | 9 | 0 | 0 |
| A6 | 16 | 635 | 0 | 0 | 13 | 495 | 0 | 0 | 1 | 26 | 0 | 0 | 2 | 114 | 0 | 0 |
| A7 | 101 | 370 | 0 | 0 | 46 | 207 | 0 | 0 | 15 | 45 | 0 | 0 | 40 | 118 | 0 | 0 |
| A8 | 110 | 371 | 24 | 2 | 76 | 293 | 13 | 2 | 13 | 35 | 4 | 0 | 21 | 43 | 7 | 0 |

表 3: 省略箇所検出の一致数とその頻度

| 一致数 | 全事例 | ワ格 | ヲ格 | 二格 |
|--------------|-------|-----|-----|-----|
| 8 | 8 | 8 | 0 | 0 |
| 7 | 41 | 39 | 0 | 2 |
| 6 | 70 | 65 | 5 | 0 |
| 5 | 112 | 97 | 7 | 8 |
| 4 | 135 | 105 | 9 | 21 |
| 3 | 211 | 150 | 17 | 44 |
| 2 | 352 | 219 | 36 | 97 |
| 1 | 1,049 | 593 | 115 | 341 |
| NAIST コーパスのみ | 599 | 504 | 78 | 17 |

表 5: 文内の省略アノテーションの分類

| カテゴリ | 頻度 |
|-------------------------------|----|
| 「A の B」/「AB」の A に先行詞となる表現が出現 | 16 |
| 括弧の外 (中) の省略が括弧の中 (外) の先行詞を指す | 14 |
| 後方照応 | 13 |
| その他 | 11 |

着目すべき点は文間の前方照応の個数が作業者横断的に他の種類と比較して多くアノテーションされている点である。これは、例えば、NAIST テキストコーパスでアノテーションされた文脈照応となる省略のうち、約 6 割が文内の照応関係となっている点と大きく異なっている。この違いの理由として、既存の省略検出の定義では述語の格要素が係り受け関係にない場合に省略とみなし、その先行詞を補完するようアノテーションさせていることが影響していると考えられ、逆にここでアノテーションされた文間の省略関係を詳しく調査することで、「自然な省略」として考慮すべき省略関係を定義できる見込みがある。そこで、次節では、文内と文間の省略関係がそれぞれどのような場合にアノテーションされたかを調査した結果について報告する。

4 アノテーションの結果の人手分析

4.1 文内の省略アノテーションの分析

まず、省略とその先行詞が同一の文内に出現している場合について、表 3 の一致数が 3 以上、つまり、3 名以上が省略を検出した事例のうち、どのアノテーション結果も文内の項を選択している事例 48 事例に対して、どのような現象が省略を認識するために影響しているのかを人手で分類し、それぞれの頻度を調査した⁷。この結果を表 5 にまとめる。表 5 に示した結果からわかるように、文内の省略が認識される特徴の一つとしては、例 (1) の『国民の「代理人」』のように「A の B」の A が先行詞として補完される場合や、「日本企業」の「日本」のような名詞連続の一部を先行詞とする場合が多いことがわかる。

(1) もし、議員を、国民の「代理人」だと (ϕ_1 ガ) 誤解したまま、この選挙制度に臨むことになれば、金権選挙、利益誘導政治は解消されるどころか、一層まん延することが目に見えている。

一方、省略の先行詞補完の有益な手がかりとなる助詞「は」で主題化された先行詞候補はここで分析した 48 事例中 3 事例しか出現しておらず、文内に「は」で主題化

された表現がすでに出現している場合にはさらに追加して補完を行うことが少ないことがわかる。

また、文を単位としている場合であっても、括弧をともなう引用等が含まれる場合には、例 (2) のように、括弧内に複数の文が含まれる場合があり、これを同一文内の省略とみなすかには議論の余地がある。

(2) デルタ農民¹が抱える課題は途方もなく大きく、「日本のような近代農業に、(ϕ_1 ガ) 太刀打ちできない。土地を手放す農民が増えるのも当然です」と同所長は嘆く。

4.2 文間の省略アノテーションの分析

次に、文間の省略を分析するが、この分析の過程で、文内に複数の述語があり、その項を他の文に出現する表現で補完する際に、アノテーション作業者は複数ある述語の一つのみを対象に省略補完し、それ以外の述語には先行詞を補完しないという傾向が見られた。そこで、NAIST テキストコーパスにアノテーションされた述語項関係の情報をもとに、文内に同一の項を持つアノテーションされている複数の述語については、その述語集合を 1 事例とみなして、述語集合の少なくとも一つについて先行詞が記入されている場合には、その事例に対して省略が補完され、そうでない場合には省略が検出されなかったという緩やかな一致の基準に基づいて、表 3 に示した一致数を再度算出した。結果を表 6 にまとめる。この結果、特にワ格に対して、一致数が増加していることがわかる。そこで、この結果を利用し、表 6 の一致数の頻度が 8 もしくは 7 の事例 69 事例⁸を対象に、どのような表現が文間の省略の先行詞となったのか等を調査し、各現象を分類してその頻度を求めた。この結果を表 7 にまとめる。表 7 に示した結果より、文間で頻出して省略と認定された事例では、述語が全て動詞であること、また、先行詞のほぼ全てが人名や組織名、人を表す普通名詞等の具体名詞であり、これに該当する典型的な事例を例 (3) に示す。

(3) カンボジアのシアヌーク国王¹は 3 1 日、中国の江沢国民国家主席と会談。新年早々、カンボジアへ (ϕ_1 ガ) 帰国する予定を (ϕ_1 ガ) 伝えた。

逆に、抽象名詞が先行詞として補完されたのは名詞句「規制緩和」の 1 件だけであった。これらの特徴については、分析対象が新聞記事であることが影響していることも考えられるため、今後の調査では多様な文書集合を対

⁷4.1、4.2、4.3 の頻度調査では一つの事例が複数のカテゴリに分類される場合は、各カテゴリに対してそれぞれ一つずつカウントした。

⁸この 69 事例はすべて文間照応の事例である。

表 6: 省略箇所検出の一致数 (緩和した一致) とその頻度

| 一致数 | 全事例 | ガ格 | ヲ格 | ニ格 |
|-----|-----|-----|-----|-----|
| 8 | 20 | 20 | 0 | 0 |
| 7 | 49 | 47 | 0 | 2 |
| 6 | 68 | 63 | 5 | 0 |
| 5 | 99 | 83 | 8 | 8 |
| 4 | 101 | 74 | 6 | 21 |
| 3 | 181 | 120 | 17 | 44 |
| 2 | 328 | 195 | 36 | 97 |
| 1 | 996 | 541 | 114 | 341 |

表 7: 文間の省略アノテーションの分類

| カテゴリ | 頻度 |
|------------------|----|
| 省略を持つ述語が動詞 | 69 |
| 先行詞が具体名詞 | 68 |
| 先行詞が抽象名詞 | 1 |
| 先行詞が「は」で主題化されている | 37 |
| 主節の動詞に対して省略を検出 | 51 |
| 1 文前に先行詞が出現 | 46 |

象にした調査が必要だと考えられる。また、文間の場合では助詞「は」で主題化された表現が先行詞として補充される場合が全体の半数以上であり、人間が文内と文間で特定している省略関係が異なっていることがわかる。また、全体の66%は1文前から先行詞が補充され、さらに、それに該当しない場合であっても、1文前の文で正しく先行詞が補充されている場合には、そこに先行詞が出現する事例が23件中12件あり、ほとんどの文間の省略補充は1文前から先行詞が補充されることがわかる。これも、文書の記述スタイルに関する影響を受けると考えられるため、他のドメインでの調査が必要だと考えられる。

4.3 NAIST テキストコーパスにのみアノテーションされた文間の省略の分析

次に、NAIST コーパスではアノテーションされているが、8名のどの作業員もアノテーションしなかった文間の省略関係を調査することで、人間が自然と思う省略と既存のコーパスに付与された省略関係のずれを考える。調査対象としている100記事のうち、この条件に合致する事例は96件(全体の約7%)存在したため、これらの事例を人手で分類し、その頻度を調査した。この調査結果を表8に示す。表8から、NAIST コーパスにのみアノテーションされた事例には、先行詞が2文以上離れた位置に出現する、また、動詞以外の述語に対して省略を検出している、先行詞として補充された表現が抽象名詞である等の特徴があることがわかり、これらは例えば、下記の例(4)や例(5)のような事例が該当する。

- (4) ... 香港のキャセイ航空や台湾の中華航空にとっては最大のドル箱路線¹だ。その大きな理由は台湾が中国との直行便運航を認めておらず、年間百数十万人もの台湾人が香港経由で中国に渡るからだ。(φ₁ ガ) 旅客機だけではない。
- (5) 挑戦者、米長邦雄の二勝一敗で迎えた一局。銀のタダ捨て¹である! 控室は騒然。「命がけで指しているものにだけ控かぶ(φ₁ ガ) 手だ」と立会だった板谷進九段がうなった。棋史に残る(φ₁ ガ) 妙手。

表8に示した特徴は表7に示した人間が文間の省略とみなした事例と逆の傾向を表しているため、今後、コンセンサスを得られるような省略解析の問題を検討する上で、これらの特徴の差異を考慮することが重要となると考えられる。

表 8: NAIST コーパスにのみ付けられた文間の省略の分類

| カテゴリ | 頻度 |
|------------------|----|
| 先行詞が2文以上離れた位置に出現 | 40 |
| 先行詞が抽象名詞 | 36 |
| アノテーション仕様のずれ | 31 |
| 動詞以外の述語の省略 | 30 |
| 節照応 | 6 |
| その他 | 15 |

5 おわりに

本研究では、これまでの既存研究で導入されてきた省略解析の問題の設定方針に対し、「人間が自然に省略と判断する省略の現象が省略解析の問題としてコンセンサスの得られる問題集合となる」という仮説に基づき、複数人の作業員が同一の文書集合に対し省略検出のアノテーションを行い、その結果を用いた人手による分析を行った。この分析の結果、文内と文間の省略の検出に異なりがあることや、省略関係がアノテーションされた既存コーパスであるNAISTテキストコーパスとどのような点が異なっているかを明らかにした。本研究は、複数人のアノテーション作業員を用いた事例調査の一例でしかないが、この取り組みの延長として、クラウドソーシングを用いた大規模な人員を導入した自然な省略の検出の事例収集や、その結果を使った分析を考えており、最終的には研究コミュニティでコンセンサスの得られる省略解析の問題の定義、省略解析の射程を策定し、文を越えた処理として対象とすべき省略解析の問題と意味役割付与のような既存の文内の解析の問題を整理し、言語処理の基盤的な解析の問題に対して、意味のあるタスク分割を提案したいと考えている。

参考文献

- [1] Iida, R. and Poesio, M.: A Cross-Lingual ILP Solution to Zero Anaphora Resolution, *Proceedings of ACL-HLT 2011*, pp. 804-813 (2011).
- [2] Imamura, K., Saito, K. and Izumi, T.: Discriminative Approach to Predicate-Argument Structure Analysis with Zero-Anaphora Resolution, *Proceedings of ACL-IJCNLP*, pp. 85-88 (2009).
- [3] Miller, G. A.: WordNet: A Lexical Database for English, *Communications of the ACM*, Vol. 38, pp. 39-41 (1995).
- [4] Mitkov, R.: *Anaphora Resolution*, Studies in Language and Linguistics, Pearson Education (2002).
- [5] Nakaiwa, H. and Ikehara, S.: Intrasentential Resolution of Japanese Zero Pronouns in a Machine Translation System using Semantic and Pragmatic Constraints, *Proceeding of the 6th TMI*, pp. 96-105 (1995).
- [6] Ng, V.: Supervised Noun Phrase Coreference Research: The First Fifteen Years, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 1396-1411 (2010).
- [7] Sasano, R., Kawahara, D. and Kurohashi, S.: A Fully-Lexicalized Probabilistic Model for Japanese Zero Anaphora Resolution, *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pp. 769-776 (2008).
- [8] 河原大輔, 黒橋禎夫, 橋田浩一: 「関係」タグ付きコーパスの作成, 言語処理学会第8回年次大会発表論文集, pp. 495-498 (2002).
- [9] 工藤拓, 市川宙, 中川哲治, 賀沢秀人: A joint inference of deep case analysis and zero subject generation for Japanese-to-English statistical machine translation, 言語処理学会第20回年次大会発表論文集, pp. 290-293 (2014).
- [10] 村田真樹, 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).
- [11] 飯田龍, 小町守, 井之上直也, 乾健太郎, 松本裕治: 述語項構造と照応関係のアノテーション: NAIST テキストコーパス構築の経験から, 自然言語処理, Vol. 17, No. 2, pp. 25-50 (2010).