

述語対の項共有情報を利用した文間ゼロ照応解析*

飯田 龍 徳永 健伸
 東京工業大学 大学院情報理工学研究所
 {ryu-i,take}@cl.cs.titech.ac.jp

1 はじめに

文章中の省略された述語の格要素を補完するゼロ照応解析は情報抽出や機械翻訳などの応用処理において必須の要素技術であり、これまでにさまざまな研究が進められてきた [10, 13, 4, 8, 12]。ゼロ照応解析の問題は省略された格要素を検出するゼロ代名詞検出と、検出されたゼロ代名詞の先行詞を先行文脈から同定する先行詞同定の 2 つの部分問題に分けて考えることができる。本研究では特に先行詞同定の問題に着目し、先行詞同定にどのような手がかりが役立つかを調査した結果について報告する。

我々はこれまでにゼロ代名詞とその先行詞が同一文に出現する文内ゼロ照応解析では文の統語的な情報が解析精度向上に貢献することを明らかにした [4]。一方、ゼロ代名詞とその先行詞が異なる文に出現する文間ゼロ照応解析では、文の統語的な位置関係を利用できないため、文内ゼロ照応解析と比較して解析精度が低い。文間ゼロ照応解析の精度を向上させるためには、談話構造中の位置関係など文章の首尾一貫性に関する手がかりを利用することが考えられるが、本研究では Schank [9] のスクリプトに代表されるような事態の遷移とその遷移の中で共有される項の情報を利用することで、文間ゼロ照応関係にあるゼロ代名詞とその先行詞を関連付けることを考える。例えば、(1)a, b の文章では述語「壊す」のガ格が省略されており、その先行詞は前文の「次郎」である。

- (1) a. 太郎が次郎_iを尋ねたとき、彼_iはラジオを修理していた。先週 (ϕ_i ガ) 落として (ϕ_i ガ) 壊したらしい。
 b. 太郎は次郎_iがラジオを修理しているときに彼_iの家を訪れた。先週 (ϕ_i ガ) 落として (ϕ_i ガ) 壊したらしい。

(1)a. では「次郎」が主題化されているのに対し、(1)b. では「太郎」が主題化されており、表現の顕現性の強さだけでは正しく先行詞を同定することができない。この問題を解決するためには、述語対「 X_i ガ壊す- X_i ガ修理する」の間の項の共有関係に関する知識を利用することが考えられる。このような述語対の知識獲得に関しては、近年含意関係認識 [1] を目的とした事態間関係知識の自動獲得の研究が盛んに行われており [7, 15, 11]、この獲得手法をもとに作成した資源をゼロ照応解析に適用することが考えられる。これらの研究では述語対があるパタンで共起する、または文章内に複数回出現する名詞句

(アンカー) を経由して共起するといった手がかりをもとに計算した述語対の共起尺度によって述語対に関係があるか否かを判断している。アンカーを經由して述語対の共起を収集することは、同一文字列の名詞句が照応関係にある場合の述語対を収集することに相当するが、文章中の照応関係は照応詞が代名詞で出現する、もしくは照応詞が省略されるなど、必ずしも同一の表現として出現するとは限らないため、アンカーを利用するだけでは文章中のすべての照応関係と関連する述語対を収集することは困難である。また、アンカーを用いて述語対を収集した場合、同一文字列の名詞句が照応関係にない場合にも述語対を抽出してしまうという欠点も存在する。さらに、同一名詞句の共参照関係は人名や組織名が多くなるといった偏りがあるため、収集できる述語対にも偏りが生じる可能性がある。

これに対し、提案手法では、タグ付与された述語項構造の情報をもとに項共有の分類モデルを作成し、そのモデルを利用して大規模コーパスから項を共有する可能性のある述語対を収集する。これにより、アンカーを用いた場合と比較してノイズの少ない述語対が収集可能になると考えられる。また、収集した述語対インスタンスからどの程度項を共有しやすいかのスコアを計算する方法についても提案を行う。

本稿では、まず 2 節で関連する事態間関係知識の獲得に関する関連研究を紹介し、3 節で提案する述語対が項を共有する事例の収集方法と収集した事例の利用方法について説明する。次に 4 節では 3 節で示した項の共有情報を文間ゼロ照応の先行詞同定に適用した結果について報告する。最後に 5 節でまとめる。

2 関連研究

含意関係認識など、言語の理解に関連するタスクで利用するためにさまざまな事態間関係の知識獲得の手法が提案されている。これらの手法の多くは述語対があるパタンで共起することを手がかりに述語対のインスタンスを収集し、その結果を自己相互情報量などの共起尺度でスコア付けすることで知識獲得を試みている。共起事例を収集するための手がかりとしては、(a) 並列関係 [15]、(b) 述語の局所文脈の類似性 [5, 11]、(c) アンカー [7] などがあるが、ここではアンカーを用いた知識獲得の一例として、Pekar ら [7] の手法を紹介する。Pekar [7] の手法では、同一文章内に出現する同一の表現 (アンカー) が係る動詞の間には何らかの関係があるという仮説に基づき知識獲得を行っている。彼らの手法では、例えば “Mary bought a house. The house belongs to Mary.” に出現している “Mary” と “house” の 2 つのアンカーを参考に、各アンカーに関連する {buy(obj:X), belong(subj:X)}

* Inter-sentential Zero-anaphora Resolution Exploiting Automatically Acquired Argument-sharing Information
 Ryu Iida, Takenobu Tokunaga
 Graduate School of Information Science and Engineering, Tokyo Institute of Technology

and {buy(subj:X), belong(to:X)} という 2 つの動詞集合を抽出し、この抽出した動詞集合の構成要素の頻度から動詞対の結び付きの強さを求めている。

我々が以前行った予備調査 [17] の結果、(a), (b), (c) の 3 種の手がかりのうち、アンカーを利用して述語対を収集した結果が最も先行詞同定に貢献したため、(c) のアンカーに基づく知識獲得の手法を、本論文で提案する手法との比較対象とする。Pekar[7] が提案した手法では、動詞対を抽出する範囲を文の距離もしくは段落の境界で制御するが、その最適な点をあらかじめ知ることは難しい。そこで、本研究では単純に同一文章内からアンカーとなる候補を抽出し¹、同一アンカーに係る動詞集合中の任意の対を動詞対の頻度として計上する。この頻度に基づき、自己相互情報量を計算することで、述語対の項共有のスコアとした。ただし、後述するように本稿では頻出するガ格のゼロ照応に着目して解析を行うため、アンカーがガ格で動詞に係る場合のみを抽出の対象とした。

3 項共有情報の自動獲得とその利用

提案手法では、以下の 2 段階の手続きで述語対がどの程度項を共有しやすいかのスコアを見積もる。

1. 述語対の格項が同一名詞句をとるか否かの分類モデル（項共有分類モデル）を作成。
2. 大規模タグ無し文章集合中の述語対を項共有分類モデルを利用して分類。この結果を利用して、述語対の項共有のスコアを見積もる。

以下の節で各手続きについて説明する。

3.1 項共有の分類モデル

まず、述語項構造の関係をタグ付与されたコーパスを利用し、述語対が同一名詞句を項として持つか否かを分類するモデルを作成する。この際、述語対が同一の名詞句を指している場合を正例、それ以外の場合を負例として学習を行う。ただし、この述語対を文章中に出現する任意の述語の組み合わせで考えた場合、ほとんどの組み合わせで項を共有していないため、大量の負例が生成されることになる。また、正例が十分に存在する場合でも二つの述語が異なる文、もしくは同一文内の場合でも統語的に離れた位置に出現している場合にはその項が同じ名詞句を指すかを分類するのが困難になると思われる。そこで、本研究では述語対に係り受け関係にある場合のみを対象に学習・分類を行う。述語対に係り受け関係にある場合とは基本的には従属節の主辞である述語が主節の述語に係る場合であり、このような関係は南 [6] が議論しているように「ため」「が」「ので」のような接続表現が項共有の手がかりとなる。また、このような手がかり表現が出現していない場合でも述語の組み合わせを訓練事例から学習することで、任意の箇所出現する述語対を対象とした場合と比較して分類が簡単になると考えられる。

また、分類対象となる格の組み合わせについては、NAIST テキストコーパス [18] 中の係り受け関係にある述語対 37,787 事例を対象に格の組み合わせのそれぞれについてどの程度同一の項を指すかの頻度を調査した。

¹品詞が“名詞-接尾-*”、“名詞-副詞可能”、“名詞-代名詞-*”、“名詞-非自立-*”、“名詞-数”、“名詞-特殊-*”、“名詞-動詞非自立的”、“名詞-接続詞的”、“名詞-引用文字列”を除くすべての名詞をアンカーの候補とした。

表 1: 格の組み合わせに関する項共有の頻度

係り元 \ 係り先	ガ格	ヲ格	ニ格
ガ格	8,543	503	222
ヲ格	638	479	79
ニ格	193	60	69

表 2: 項共有の学習・分類に利用する素性

素性	説明
PRED_WORD _{i,j}	係り元の述語 v_i (係り先の述語 v_j) の述語の文字列
FUNC_IN_PRED _{i,j}	係り元の述語 v_i (係り先の述語 v_j) を含む文節に出現する機能語の文字列
FUNC_BTWN_PREDS	係り元の述語 v_i と係り先の述語 v_j の間に位置する文節に出現する機能語の文字列
FUNC_IN_DAUGHTERS _{i,j}	係り元の述語 v_i (係り先の述語 v_j) の係り先の文節に出現する機能語の文字列
JS_DIVERGENCE	係り元の述語 v_i の $p(z v_i)$ と係り先の述語 v_j の $p(z v_j)$ の JS divergence

この結果を表 1 に示す。表 1 からわかるように、格の組み合わせによっては分類のための訓練事例が十分にあるとはいえない。そこで、本研究では係り元と係り先の述語がともにガ格で項を共有する場合に限定し、教師ありの項共有モデルを構築する。

学習・分類には表 2 に示した素性を利用する。表 2 のうち、特に FUNC_IN_PRED_i はどのような接続表現を伴って係り先の述語に係るかの特徴を捉える。また、FUNC_IN_DAUGHTERS_{i,j} は述語 v_i と v_j がどのような格要素を伴って出現しているかを表す。この素性は、例えば「太郎_i は学校へ (ϕ_i ガ) 行って授業を受けた」のように係り元もしくは係り先の述語のガ格が明示的に出現していない場合は項を共有している可能性があり、一方「太郎は学校へ行き、次郎は病院に行った」のように二つの述語がガ格に相当する項を取る場合は項を共有する可能性が低くなるという特徴を捉える。また、訓練事例中に出現する述語対だけでなく、他の大規模コーパスから収集した述語対の情報も分類の際の手がかりになると考えられる。ここでは、毎日新聞 1991 年から 1994 年、1996 年から 2003 年までの記事全体を係り受けした結果から〈名詞 n 、格助詞 c 、動詞 v 〉の係り受け関係を収集し、これを用いてある動詞 v_i が別の動詞 v_j とどの程度同じ格要素を共有するかを計算した結果を素性として利用する。ただし、単純に共起頻度を利用した場合偶然新聞記事中で共起しない場合について適切な割合を算出できない可能性がある。そこで、本手法では、 n と c の対と v の共起行列を pLSI [3] を用いて次元圧縮し、その結果得られた隠れクラス z への帰属確率分布 $p(z|v)$ を項共有の手がかりとして使用する。この $p(z|v)$ が類似する述語は類似する項を取るため、その分布間の距離を Jensen-Shannon divergence で計算した値を素性の一つ (表 2 の JS_DIVERGENCE) として利用した。

3.2 述語対インスタンスの項共有のスコア

3.1 で作成した項共有分類モデルを大規模な係り受け解析済みのコーパスに適用することで、コーパス中に出現する係り受け関係にある述語対がガ格を共有するか否かを分類した結果を得る。この結果、ある述語対に関して複数の分類結果を得ることになる。ここで分類器はある文脈に出現する各述語対に関してどの程度項を共有しやすいかの値を出力するが、項共有分類モデルが正しく分類できると仮定すると、このモデルが出力する値が高

いほど、その述語対は項を共有しやすく、また述語対が偶然1つだけ高い値を持つよりも複数回出現した述語対がそれぞれの文脈で項を共有しやすいと分類されたほうが信頼できる手がかりになると考えられる。そこで、この2点を考慮し、最終的に式(1)に従って述語対の項共有のスコアを推定する。

$$\text{arg_share}(v_i, v_j) = \log(N_{ij} + 1) \sum_k^{N_{ij}} \text{score}(v_{ik}, v_{jk}) \quad (1)$$

ここで、 v_i, v_j はそれぞれ係り元、係り先の述語を表す。また、 N_{ij} はその述語対がコーパス中に出現した頻度を表す。 $v_{ik} (v_{jk})$ はある文脈 k に出現した係り元 (係り先) の述語を表し、 $\text{score}(v_{ik}, v_{jk})$ は v_{ik}, v_{jk} の述語対を項共有分類モデルで分類した結果出力される値である。

4 評価実験

提案手法の有効性を調査するために、述語対の知識のみで先行詞候補を順序付けした場合と他の情報も利用した教師有り手法を用いた場合の2通りの評価を行った。各結果を以下に示す。

4.1 項共有情報を利用した先行詞候補の順序付け

3節に示した収集方法を用い、同一名詞句と照応関係となる述語対を収集する。動詞対の共起頻度を抽出するために本研究では毎日新聞の1991年から1994年、1996年から2003年までの新聞記事全体を利用した。これらの記事を茶筌²とCaboCha³を用いて形態素・係り受け解析を行い、得られた結果から項共有のスコアを計算する。また、学習にはSupport Vector Machine⁴を用い、線形カーネル、パラメタ c はデフォルト値を使用した。

実験では、提案する項共有のスコアの有効性を調査するため、以前の我々の予備調査の結果、最も解析に貢献したアンカーに基づく項共有のスコアを比較対象とする。また、ベースラインとして先行詞候補集合からランダムに1つを先行詞として選択した場合の結果も掲載する。この実験では各スコアを利用することで、先行詞をどの程度上位に順位付けできるかを式(2)を用いて評価する。

$$MRR = 1/|N| * \sum_{n \in N} 1/\text{rank}(n) \quad (2)$$

ここで N は事例の数を表し、 $\text{rank}(n)$ は事例 n 中に出現する先行詞候補を各スコアで順序付けした場合の先行詞の順位を表す。

評価には、NAIST テキストコーパスの一部 [18] から287記事 (合計2399文) にタグ付与された文間ゼロ照応の関係のうち頻出するガ格でかつ先行詞が述語に係る場合394事例を利用する。それぞれの手法について式(2)のMRRの値を求めた結果を表3に示す。表3より、提案手法が最も高いMRR値を得ているため、提案手法のスコアを先行詞同定に導入することにより、解析精度がより向上する可能性がある。これについては4.2で比較実験を行う。また、アンカーを用いた手法では1,651,118対の述語対インスタンスを収集できたのに対し、提案手法では9,146,543対もの述語対インスタンスを収集でき

²<http://chasen.naist.jp/hiki/ChaSen/>

³<http://chasen.org/~taku/software/cabocho/>

⁴<http://svmlight.joachims.org/>

表3: 各モデルの平均MRR

モデル	MRR	NA
ベースライン	0.436	-
アンカーを用いた手法	0.465	53
提案手法	0.524	15

NAはそれぞれの手法で項共有のスコアを求めることができなかった事例の頻度を表す。

表4: 各手法で獲得された述語対と項共有に関する信頼度

述語対	アンカーに基づく手法		提案手法	
	スコア	述語対	スコア	述語対
適合する-適合する	9.49	逮捕する-捜索する	5.96	
亡くす-亡くす	8.72	語る-示す	3.71	
着陸する-着陸する	7.75	置く-始める	3.40	
見守る-すすり泣く	7.17	語る-表明する	3.32	
告訴する-告訴する	6.95	語る-見せる	2.89	
引き下げる-引き下げる	6.83	示す-けん制する	2.61	
造る-造る	6.28	亡くす-する	2.55	
溶ける-溶かす	6.12	亡くす-失う	2.46	
捨てる-捨てる	6.07	訴える-示す	2.35	
収容する-収容する	5.78	強調する-示す	2.05	
重ねる-かまう	5.67	策定する-目指す	2.01	
捜索する-送検する	5.59	示す-示す	1.96	
送検する-捜索する	5.59	主張する-見せる	1.89	
起訴する-起訴する	5.53	導入する-高める	1.86	
刊行する-創刊する	5.52	強調する-呼びかける	1.79	
下りる-下山する	5.51	求める-調べる	1.76	
追う-追う	5.43	あいさつする-示す	1.76	
検証する-捜索する	5.38	見直す-作成する	1.74	
導入する-導入する	5.27	逮捕する-調べる	1.71	
作成する-作成する	5.12	強調する-表明する	1.68	

ている。この結果、アンカーを用いた場合に比べて提案手法では項共有のスコアを計算できる述語対の数が多くなるという結果も得られた。

それぞれの手法で得られた述語対の上位20件を表4示す。この表からわかるように、アンカーを用いた場合では同一述語の対が高い値を得ているのに対し、提案手法のスコアが高い述語対はアンカーを用いた場合と比較してよりスクリプト知識のような何らかの事態間関係を想起できるものとなっている。

次にタグ無しデータを増減させた場合の提案手法のMRRの変動を調査した。結果を図1に示す。この結果より、抽出したデータの量を増やすにしたがい、MRRの値が増加しており、さらにデータを増やすことでさらにMRRの値が向上する見込みがある。今回の手法では自動解析した係り受けの関係と、少量の述語項構造タグ付きのデータから学習した項共有のモデルを利用しているだけであり、大規模なタグ無し文章集合を用意するだけで容易にデータの規模を増加させることが可能である。ただし、本研究では記事集合として新聞記事を利用しているため、Webページなど他の記事集合で同様の数値が得られるとは限らない。これについても今後調査を進めたい。

4.2 文間ゼロ照応解析の精度比較

次に項共有の情報が文間ゼロ照応解析の精度にどのように影響するかを調査するために、提案する項共有のスコアやアンカーに基づくスコアを機械学習に基づく先行詞同定モデルの素性として利用した場合の精度の変化について調査した。

4.2.1 先行詞同定モデル

先行詞同定には我々が以前提案したトーナメントモデル [16] を利用する。トーナメントモデルでは、先行詞候補の対でどちらがより先行詞らしいかを判定する分類モデルを利用し、2つの候補間の勝ち抜き戦を行うこと

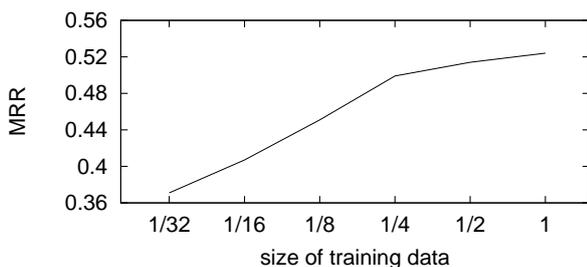


図 1: 提案手法の学習曲線

表 5: 文間ゼロ代名詞の先行詞同定の精度

モデル	精度
a) ベースラインモデル	0.668
b) a)+ANCHOR	0.689
c) a)+ARG_SHARE	0.696
d) a)+ANCHOR,+ARG_SHARE	0.705

ANCHOR: アンカーに基づく項共有のスコア,
ARG_SHARE: 提案手法の項共有のスコア.

により, 最も先行詞らしい候補を決定する. 候補間の比較には以下の素性を利用する.

- 対象となるゼロ代名詞を項として持つ述語の語彙・統語情報に関する素性.
- 先行詞候補に関する語彙・統語・意味情報, 出現位置に関する素性. また, ゼロ代名詞が出現した状況におけるセンタリング理論に基づく先行詞候補の顕現性の情報.
- ゼロ代名詞を項として持つ述語と先行詞候補の対から抽出可能な素性(選択選好や述語と先行詞候補の距離など).

トーナメントモデルや素性の詳細については文献[16]を参照されたい.

4.2.2 先行詞同定の評価結果

4.2.1 で示した素性に加え, アンカーに基づく項共有のスコアや提案する項共有のスコアを素性に加えた場合の先行詞同定の精度の変化を 4.1 で使用した評価データを対象に調査を行った. 結果を表 5 に示す. 表 5 より, アンカーに基づく項共有のスコア, 提案手法の項共有のスコア, それぞれを加えた場合にベースラインと比較して精度が向上していることがわかる. 特に, 後者の提案手法のスコアを用いた場合により精度が向上している. また, 両方を素性として利用した場合にはさらに精度が向上するという結果が得られた. つまり, 異なる方法で収集された述語対の情報を組み合わせることにより精度が向上しており, この点については各スコアを求めるために利用したコーパスの規模との関連性も含め, 今後さらに調査する予定である.

5 おわりに

本稿では, ゼロ照応関係タグ付きコーパスを用いて項共有分類モデルを作成し, そのモデルを利用して述語対の項共有関係のスコアを推定する手法を提案した. また, このスコアを機械学習に基づくゼロ照応解析モデルの素性として利用することで文間ゼロ照応解析の先行詞同定の精度を向上させることができた.

提案手法では, 共起が疎になるのを回避するために述語単体の対の情報を利用したが, 実際は“体調ヲ崩す → 病氣ニなる”のような句と句の粒度で事態間の関係を捉えることが望ましい. 句の意味表現に関しては, 語の共

起ベクトルにベクトル和や積などの何らかの演算を行い句の意味を生成的に求める手法が提案されており [2, 14], これらの手法と本手法を統合することで, 句レベルの項共有の資源を構築できる可能性がある. また, 含意関係認識の資源構築では, アンカーを用いた獲得手法のように項共有情報が重要な手がかりとなる. そのため, 本稿で導入した項共有分類モデルを利用することで含意関係認識の精度向上につながる資源の作成が可能になる見込みがある. これら 2 点を含め, 提案手法の応用については今後の課題としたい.

参考文献

- [1] Dagan, I., Glickman, O. and Magnini, B.: The PASCAL Recognising Textual Entailment Challenge, *Machine Learning Challenges. Lecture Notes in Computer Science*, Vol. 3944, pp. 177–190 (2006).
- [2] Erk, K. and Padó, S.: A Structured Vector Space Model for Word Meaning in Context, *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 897–906 (2008).
- [3] Hoffman, T.: Probabilistic latent semantic indexing, *Proceedings of ACM SIGIR*, pp. 50–57 (1999).
- [4] Iida, R., Inui, K. and Matsumoto, Y.: Exploiting Syntactic Patterns as Clues in Zero-Anaphora Resolution, *Processing of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING/ACL)*, pp. 625–632 (2006).
- [5] Lin, D. and Pantel, P.: Discovery of inference rules for question answering, *Natural Language Engineering*, Vol. 7, No. 4, pp. 343–360 (2001).
- [6] 南不二男: 現代日本語の構造, 大修館 (1974).
- [7] Pekar, V.: Acquisition of verb entailment from text, *Proceedings of Human Language Technology Conference/North American chapter of the Association for Computational Linguistics annual meeting (HLT-NAACL06)*, pp. 49–56 (2006).
- [8] 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)*, pp. 769–776 (2008).
- [9] Schank, R. and Abelson, R.: *Scripts Plans Goals and Understanding: An Inquiry into Human Knowledge Structures*, Lawrence Erlbaum Associates, Publishers (1977).
- [10] Seki, K., Fujii, A. and Ishikawa, T.: A Probabilistic Method for Analyzing Japanese Anaphora Integrating Zero Pronoun Detection and Resolution, *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, pp. 911–917 (2002).
- [11] Szepietor, L. and Dagan, I.: Learning Entailment Rules for Unary Templates, *Proceedings of The 22nd International Conference on Computational Linguistics (COLING)*, pp. 849–856 (2008).
- [12] Taira, H., Fujita, S. and Nagata, M.: A Japanese Predicate Argument Structure Analysis using Decision Lists, *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 523–532 (2008).
- [13] 磯崎秀樹, 賀沢秀人, 平尾努: 辞書式順序を持つペナルティによるゼロ代名詞解消, *情報処理学会論文誌*, Vol. 47, No. 7, pp. 2279–2294 (2006).
- [14] 谷塚太一, 飯田龍, 徳永健伸: 格要素間の依存関係を考慮した選択選好モデル, *情報処理学会 自然言語処理研究会報告 NL-193* (2009).
- [15] 鳥澤健太郎: 「常識的」推論規則のコーパスからの自動抽出, *言語処理学会第 9 回年次大会発表論文集*, pp. 318–321 (2003).
- [16] 飯田龍, 乾健太郎, 松本裕治: 文脈の手がかりを考慮した機械学習による日本語ゼロ代名詞の先行詞同定, *情報処理学会論文誌*, Vol. 45, No. 3, pp. 906–918 (2004).
- [17] 飯田龍, 乾健太郎, 松本裕治: 結実性と首尾一貫性から見たゼロ照応解析, *情報処理学会 自然言語処理研究会報告 NL-178*, pp. 45–52 (2008).
- [18] 飯田龍, 小町守, 乾健太郎, 松本裕治: NAIST テキストコーパス: 述語項構造と共参照関係のアノテーション, *情報処理学会研究報告 (自然言語処理研究会) NL-177-10*, pp. 71–78 (2007).