日本語書き言葉を対象とした述語項構造と共参照関係のアノテーション: NAIST テキストコーパス開発の経験から*

飯田 龍　小町 守　乾 健太郎　松本 裕治
奈良先端科学技術大学院大学
{ryu-i,mamoru-k,inui,matsu}@is.naist.jp

1 はじめに

情報抽出や機械翻訳などの NLP の応用処理への需要が高まる中で，その技術を実現するための中核的な要素技術となる共参照解析や述語項構造解析の問題に対してさまざまな手法が提案されている．それらの手法では各情報が付与されたコーパス（以後，タグ付与コーパス）を訓練用データとして教師あり手法を用いるやり方が一般的であり，解析の対象となるコーパス作成の方法論についても議論がなされてきた [3, 5, 1].

共参照解析については，主に英語を対象にいくつかのタグ付与のスキーマが提案されており，実際にそのスキーマに従ったコーパスが作成されている [3, 9, 2, 8, 1]．例えば，Message Understanding Conference (MUC) の Coreference (CO) タスク [3] や，その後継にあたる Automatic Content Extraction (ACE) プログラム [1] の Entity Detection and Tracking (EDT) タスクでは，数年に渡って主に英語を対象に詳細な仕様が設計されてきた．また，述語項構造解析に関しては，CoNLL の shared task[1] で評価データとして利用されている PropBank[7] を対象に仕様が模索されてきた．

日本語を対象に述語項構造と共参照の研究をするにあたり，分析，学習，評価のための大規模なタグ付きコーパスが必要となるが，現状で利用可能な Global Document Annotation (GDA) [2] タグ付与コーパス（以後，GDA コーパス）や京都テキストコーパス第 4.0 版（以後，京都コーパス 4.0）[9] は，述語項構造や共参照の解析のための十分な規模の評価データとはいえない．

タグ付与の仕様についても，MUC や ACE など仕様が実際に応用処理で必要とされる情報を考慮できているのか，また英語と日本語の言語の差異によって生じる問題のずれなどを考える必要がある．そこで，本稿では，タグ付与に関する既存の研究を吟味し，述語項構造と共参照情報の書き言葉コーパスへのタグ付与に関してどのような仕様を採用するかについて述べる．2 節で照応と共参照の関係について確認し，3 節で先行研究を踏まえた上で我々のタグ付与の指針を示す．4 節で現状のタグ付与の問題点を述べ，その問題についての議論を行い，最後に 5 節でまとめる．

なお，今回の作業の結果作成された述語項構造と共参照タグ付与コーパスは NAIST テキストコーパスとして公開している[2].

2 照応と共参照

照応とはある表現が同一文章内の他の表現を指す機能をいい，指す側の表現を**照応詞**，指される側の表現を**先行詞**という．これに対し，二つ（もしくはそれ以上）の表現が現実世界（もしくは仮想世界）において同一の実体を指している場合には**共参照**（もしくは**同一指示**）の関係にあるという．先行詞となる表現が固有表現になる場合など，多くの場合は照応関係かつ共参照の関係が成り立つ．例えば，文章 (1) では，代名詞 “彼 $_i$” が “村山首相 $_i$” を指しており，かつ同一の人物を指しているため，照応関係かつ共参照関係であるとみなすことができる．

(1) 村山首相$_i$ は...。彼$_i$ は ...。

これに対し，文章 (2) では，2 文目の “それ $_i$” は 1 文目の “iPod$_i$” を指しているため照応関係となるが，同じ実体を指していないため共参照関係とはならない．

(2) 太郎は iPod$_i$ を買った。次郎も それ$_i$ を買った。

このように照応関係にある場合でも，同一の実体を指している場合とそれ以外の場合が存在する．文献 [6] では，前者のような共参照かつ照応関係となる関係を identity-of-reference anaphora（IRA），後者を identity-of-sense anaphora（ISA）と呼び区別している．

照応と共参照は，IRA が両方の性質を兼ねるため，本来異なる概念であるにもかかわらず混同して扱われてきた．タグ付与コーパス作成に関する先行研究でも同様にいくつかの異なる解釈で仕様が設計されている．

3 NAIST テキストコーパスの仕様

述語項構造と共参照関係のタグ付与に関する先行研究を踏まえ，今回の作業ではおおきく (1) 述語の基本形とその表層格，(2) 事態性名詞とその表層格，(3) IRA の関係のみを対象とした共参照関係の 3 つの関係を対象にタグを付与した[3].

3.1 述語と項のタグ付与

述語そのものの認定に関しては，品詞体系として IPADIC[4]を採用し，動詞，形容詞，名詞+“だ（助動詞）” の 3 種類をタグ付けの対象となる述語とみなし，作業を行う．

述語と項の関係については，京都コーパス 4.0 が採用しているような表層格，GDA のような深層格，また PropBank で付与されているような独自の基準など，さまざまなタグ付与のレベルが考えられる．この中で我々

*Annotating Predicate-argument and Coreference Relations in Japanese Written Text: From the Experience of Building the NAIST Text Corpus
Ryu Iida, Mamoru Komachi, Kentaro Inui, and Yuji Matsumoto
Nara Institute of Science and Technology

[1] http://www.lsi.upc.edu/~srlconll/
[2] http://cl.naist.jp/nldata/corpus/
[3] 今回紙面の都合上述べることできなかった先行研究との比較については文献 [10] を参照されたい．
[4] http://chasen.naist.jp/stable/ipadic/

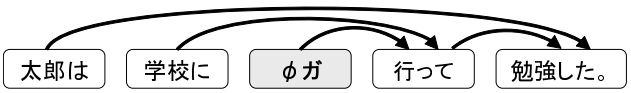

図 1: 文内ゼロ照応の認定

表 1: 述語と項のタグ付与の比較

コーパス	付与の対象	付与の範囲
PropBank	意味役割相当	intra
GDA	意味役割	inter, exo
京都コーパス 4.0	表層格（出現形）	intra, inter, exo
NAIST コーパス	表層格（基本形）	intra, inter, exo

intra:文内照応，inter:文間照応，exo:外界照応

は「誰が何を何に対してどうする」という情報抽出的な観点でタグを付与することが自然だと考え，述語の原形に対して項のタグを付与する．ただし，現状では役割レベルの記述が応用処理に寄与するのかが明らかでないため，表層格レベルでタグ付与を行う．例えば，京都コーパス 4.0 では文 (3) の述語 “食べさせる” に対して “私 $_i$”, “彼 $_j$”, “リンゴ $_k$” をそれぞれガ，ヲ，ニ格でタグ付与するのに対し，我々の仕様では述語の原形 “食べる” に対して “彼 $_j$ ガリンゴ $_k$ ヲ食べる” というタグを付与する．ただし，述語の原形に対してタグを付与する場合には使役者に相当する “私 $_i$” と述語 “食べる” の間の関係にタグが付与されないことになる．これを回避するため，格要素を増やす助動詞に対してタグ “追加ガ（ニ）格” を付与した．例えば，文 (3) では，助動詞 “させる” に対し “私 $_i$” を追加ガ格でタグ付与し，文 (4) では助動詞 “やる” に対し “彼 $_j$” を追加ニ格でタグ付与する．

(3) a. **私**$_i$ は**彼**$_j$ に**リンゴ**$_k$ を 食べさせる$_{ガ:i,\ ヲ:k,\ ニ:j}$

b. **私**$_i$ は**彼**$_j$ に**リンゴ**$_k$ を 食べ$_{ガ:j,\ ヲ:k}$させる$_{追加ガ格:i}$

(4) **私**$_i$ は**彼**$_j$ に**本**$_k$ を 読ん$_{ガ:i,\ ヲ:k}$で やる$_{追加ニ格:j}$

また，京都コーパス 4.0 では表層格を網羅する形で作業が進められたが，今回の作業では，頻出するガ/ヲ/ニ格のみを対象に作業を進め，どの程度の品質で作業ができるかを調査した．

表層格は項が係り受け関係にある場合に加え，省略がある場合にも区別せずに作業を行う．例えば，図 1 では述語 “行っ（て）” は係り関係にある文節にガ格に相当するものがないため，項が省略されているとみなすが[5]，この場合にも係り受け関係にある場合と区別せずに “太郎” にガ格のタグを付与する[6]．当面このような基準で作業を進めることで，今後深層格の情報をタグとして付与する必要がでてきた場合にも，今回の仕様で付与される情報は，agent，theme のような意味役割を付与する場合や語彙概念構造（LCS）[4] の意味述語の情報を付与する際の手がかりとして利用できると考えられる．

述語に関する我々の仕様と他のコーパスの仕様を比較すると表 1 のようになる．

3.2 事態性名詞と項のタグ付与

動詞や形容詞などの述語に加え，サ変名詞や名詞化された和語動詞が事態としての意味を伴う場合に，その名詞を**事態性名詞**と認定し[7]，述語と同様に必須格となる表層ガ/ヲ/ニ格を付与する．作業者は与えられた名詞（主にサ変名詞）が事態を表しているか否かを判定し，事態性名詞と判断した名詞（句）に対して必須格を付与する．例えば，文 (5) で出現している二つの “電話” という名詞のうち，“電話 $_i$” が「電話する」というコトを表しているのに対し，“電話 $_j$” は「（携帯）電話」というモノを表している．この状況で作業者は “電話 $_i$” のみを事態性名詞と認定し，これに対して “彼 $_a$” をガ格，“私 $_b$” をニ格として付与しなければならない．

(5) **彼**$_a$ からの 電話$_{i_{(ガ:a,\ ニ:b)}}$ によると、**私**$_b$ は彼の家に 電話$_j$ を忘れたらしい。

また，タグ付与の対象が複合語の場合は構成素を分解し，それぞれに対して事態性判別の作業を行う．例えば，「紛争仲裁」は構成素 “紛争” と “仲裁” のそれぞれの意味を構成的に組み合せてできた複合語だと解釈できるので，“紛争” と “仲裁” をそれぞれ事態性名詞と判断する．一方,「フランス革命」のような分解すると複合語の持つ意味が欠落する場合にはそれ以上分解せず，“革命” に対して事態性名詞のタグを付与しない．

3.3 名詞句間の共参照関係のタグ付与

共参照のタグ付与では，2 節で述べた IRA に加え ISA の関係も含めてタグを付与するか否かの選択肢があるが，ISA の関係まで含めてしまうと，総称名詞間の包含関係のような複雑な関係を考慮して作業を行う必要がある．例えば，文章 (6) で “本 $_i$” と “本 $_j$” はともに総称名詞であるが，“本 $_i$” が「本を意味する類に属するすべての要素」を指すのに対し，“本 $_j$” は「図書館の本（図書館に置いてある本）」を指し，二つの総称名詞の間には “本 $_i \supset$ 本 $_j$” という概念間の包含関係が成り立つ．また，二つの総称名詞 “本” と “書物” の間にも概念的な包含関係があるが，そのような包含関係を考慮しながら共参照関係を認定することは困難であり，タグの揺れの原因となる．

(6) 本$_i$ は、書物の一種で、書籍・雑誌などの印刷・製本された出版物を指す。図書館の 本$_j$ は借りることができる。

そこで，述語/事態性名詞と項の関係については ISA も含めた関係にタグ付与するのに対し，共参照に関しては IRA の関係にのみタグを付与する．ただし，ACE EDT の仕様のように，実体が組織名や場所名など数種の固有表現に限定して共参照関係のタグを付与することは，さまざまな応用分野で必要となる共参照の表現を網羅できないため望ましくない．そこで，今回の作業では，作業者にはいくつか作業の具体例とともに以下の 3 つの基準を提示するだけで，表現のクラスを限定せずに共参照関係のタグ付与を行ってもらい，どのような問題が生じるのかを調査した．

1. 照応詞は文節の主辞（最右の名詞自立語）のみに限定する．
2. 談話内に出現した名詞句のみを先行詞とする．
3. 総称名詞は照応詞，先行詞とみなさない．

既存の共参照関係のタグ付与の研究と比較すると表 2 のようになる．

[5] この例は並列表現はゼロ照応と区別すべきという議論もあるが，項が係り受け関係にない場合は統一的にゼロ照応とみなすほうが機械処理を行う際には見通しがよいと考えている．

[6] 最終的に付与される情報には係り受け関係は含まないが，京都コーパスの係り受け情報と統合することによりゼロ照応か否かの判別が可能である．表 3 でまとめる統計量も京都コーパスと統合した結果を用いて求めた．

[7] 今回の作業では「運動会」や「雨」のような表現を事態性名詞として認定しない

表 3: 述語項構造に関するタグの統計

	出現箇所	ガ格		ヲ格		ニ格	
述語	同一文節内	177	(0.002)	60	(0.001)	591	(0.027)
106,628	係り関係	44,402	(0.419)	35,882	(0.835)	18,912	(0.879)
	ゼロ照応 (文内)	32,270	(0.305)	5,625	(0.131)	1,417	(0.066)
	ゼロ照応 (文間)	13,181	(0.124)	1,307	(0.030)	542	(0.025)
	ゼロ照応 (文章外)	15,885	(0.150)	96	(0.002)	45	(0.002)
	全体	105,915	(1.000)	42,970	(1.000)	21,507	(1.000)
事態性名詞	同一文節内	2,195	(0.077)	5,574	(0.506)	846	(0.436)
28,569	係り関係	4,332	(0.152)	2,890	(0.263)	298	(0.154)
	ゼロ照応 (文内)	9,222	(0.324)	1,645	(0.149)	586	(0.302)
	ゼロ照応 (文間)	5,190	(0.183)	854	(0.078)	201	(0.104)
	ゼロ照応 (文章外)	7,525	(0.264)	42	(0.004)	10	(0.005)
	全体	28,464	(1.000)	11,005	(1.000)	1,941	(1.000)

表 2: 共参照タグ付与の差異

コーパス	特徴
GDA	IRA と ISA の関係両方に付与.
ACE EDT	IRA の関係にのみ付与．ただし，実体がいくつかの固有表現のクラスに限定されている．
京都コーパス 4.0	IRA と ISA の関係両方に付与.
NAIST コーパス	IRA の関係にのみ付与.

3.4 統計

3.1， 3.2， 3.3 の仕様に従い，京都コーパス 3.0 の全記事（2,929 記事，38,384 文）を対象に，2 人の作業者が述語項構造と共参照の関係についてタグ付与作業を行った．述語/事態性名詞とその項に付与されたタグの個数を表 3 にまとめる．ただし，項の出現位置によって，同一文節内[8]，係り関係にある場合[9]，文内のゼロ照応関係，文境界を越えるゼロ照応（文間ゼロ照応）関係，文章内に項が出現しない文章外ゼロ照応の 5 つに分類して頻度を求めた．表 3 より，述語の項目ではヲ格，ニ格の項のほとんどは係り関係にあるのに対し，ガ格の約 6 割はゼロ照応の関係にあることがわかる．これに対して，事態性名詞のヲ格，ニ格は同一文節内，つまり複合語の構成素として項が出現している割合が高く，ガ格に関しては約 8 割がゼロ照応の関係にあり，述語の場合と比較して項の出現箇所がおおきく異なっていることがわかる．

共参照関係のタグについては，タグ付与された実体の総数が 10,531，最初に出現した表現を先行詞，その他を照応詞とみなしたときの照応詞の個数が 25,357 であった．京都コーパス 4.0 より圧倒的に個数が少ないが，これは実体間の関係にのみ限定して作業を行ったためだと考えられる．

次に，実際に作業を行っている 2 人の作業者間のタグ付与の一致率を調査するため，ランダムに選択した報道 30 記事を対象に作業を行った．評価は一方の作業者のタグ付与の結果を正解，他方の作業結果をシステムの出力とみなし再現率と精度で評価する．ただし，それぞれのタグの一致率は各タグの終了位置の一致で評価した．また，述語と事態性名詞の項の一致率については，2 人の作業者の述語（事態性名詞）が一致した箇所のみを対象に評価した．これらの基準で評価した結果を表 4 に示す．表 4 よりわかるように，それぞれのタグ付与は多くの場合 8 割を越える品質で作業ができているが，改善の余地は大きい．

[8]「明らかになる」のような表現がコーパス中では一文節であるのに対し，作業者が「明らかに」と「なる」を分けて付与した場合などを含む.

[9]「サンマを焼く男」の「男」が「焼く」のガ格となるような，連体修飾の関係も含む.

表 4: タグの一致率

	再現率		精度	
述語	0.921	(806/875)	0.944	(806/854)
ガ格	0.823	(683/830)	0.829	(683/824)
ヲ格	0.899	(329/366)	0.954	(329/345)
ニ格	0.724	(105/145)	0.890	(105/118)
事態性名詞	0.965	(247/256)	0.792	(247/312)
ガ格	0.735	(191/260)	0.743	(191/257)
ヲ格	0.827	(86/104)	0.869	(86/99)
ニ格	0.389	(7/18)	0.583	(7/12)
共参照	0.813	(126/155)	0.813	(126/155)

4 タグ付与の問題点と今後の展望

この節では， 3 節で述べたタグ付与作業で生じた主要な問題を説明し，その問題を解決するための今後の方向性について議論する．

4.1 事態性名詞タグ付与の問題点

事態性名詞の認定に関しては，述語の場合とは異なり，対象となる名詞（句）が出現文脈でモノとコトのどちらを表しているかを認定する作業が必要となるが，これに加え,「投資 率」のような複合語に関してはどの程度構成的に分解できるかを判断しなければならず，この分解についての基準が作業者間で異なったために一致率が低下した．また，“契約”，“規制”，“投資” のような表現は事態として解釈可能であるが，文脈によってコトを表しているのか，事態が起こった結果できた結果物としてのモノなのかの判定が困難な場合が存在した．例えば，文 (7) では “インセンティブ規制” を結果物とみなすか，“規制” を事態とみなすかで揺れが生じた．

(7) 料金規制当局と公益事業者が、一種の社会契約を結んだという考えに立つもので、経営効率化促進のための社会契約インセンティブ 規制 とも言われる。

また，現状では主にサ変名詞を対象に，ガ/ヲ/ニ格の表層格で項との関係を付与しているが，事態性名詞は述語と異なり，助詞 “を” で述語に係っている格要素が基本的にはヲ格となるといった表層格による統語的な制約を受けない．そのため，表層格でタグを付与することが必ずしもよいというわけではなく，今後 “運動会” のような広い意味での事態をサ変名詞で表現される事態と同じ枠組みで扱うことを考えた場合，各事態について agent, theme のような役割レベルでタグ付与するという選択肢もあり，どのように仕様を設計するかは今後検討したい．

4.2 項のタグ付与の問題点

項のタグ付与に関しては，述語が取り得る格パタンが複数存在するために作業者間で揺れが生じることがわかった．この問題の典型的な例が自動詞と他動詞の交替である．例えば，述語 “実現する” は同じ語義に対して表

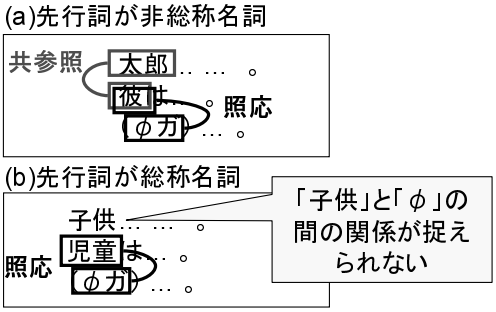

図 2: 先行詞が総称名詞の場合のタグ付与の漏れ

層格レベルで “agent ガ theme ヲ実現する” と “theme ガ実現する” の 2 つの格パタンが存在するため，述語のすべての格要素が省略されている場合は，作業者はどちらの解釈でもタグ付与が可能になってしまう．自他交替の問題と類似して，制度などの表現に動作主性（agentivity）を認めるか否かで解釈が異なるために揺れが生じる場合もある．例えば，文 (8) において，述語 “しばる” は，直前に出現している “規制” の動作主性を認めた場合，“規制ガ theme ヲしばる” と “agent ガ規制デ theme ヲしばる” の 2 つの解釈が存在する．

(8) 我々の生活が知らず知らずにどれだけ規制で しばられているか、規制緩和によって豊かさが変わっていくのかを考えてみた。

このような交替を伴う場合の揺れに関しては，どちらかのパタンを優先するという規則をあらかじめ決めて作業することで対応できると考えられる．

また項同定の問題において，図 2(a) のように，項としてタグ付与されるべき名詞句が IRA の関係で他の名詞句と関連付けられている場合，共参照関係にある名詞句のいずれかを項として同定する問題とみなすことができるが，一方図 2(b) の “子供” と “児童” のような ISA の関係で出現している名詞句については，ラベルが付与されていない “子供” は述語の項としてタグが付与されないという問題が起こる．

4.3 共参照タグ付与の問題点

IRA のみを対象に共参照のタグを付与する作業に関してもいくつかの問題が残る．例えば，IRA の認定に関して，実体が具体名詞である場合は 2 つの言及が同一の実体を指すか否かの認定が容易であるが，抽象名詞の場合は同じものを指しているかの判定が困難である． 3.3 で共参照関係のタグ付与にはあらかじめ名詞のクラスを指定して作業を行うことは望ましくないと述べたが，抽象名詞に関してはいくつかの意味クラスに限定して作業を行い，どのくらい揺れなく作業できるかを調査したい．

5 おわりに

本稿では，日本語を対象とした述語項構造・共参照タグ付与コーパスに関して，我々が今回採用したタグ付与の基準について報告した． 3 節の議論に基づき，述語項構造のタグに関しては ISA と IRA の関係両方で，共参照関係は IRA の関係でタグ付与作業を行い，京都コーパス 3.0 を対象にこれまでにない大規模な述語項構造・共参照タグ付きコーパスを作成した．また，作業の過程で起こった問題について考察し，作業の詳細化のための項目を述べた．

今後の課題としてさまざまな問題が残るが，その中の一つに任意格を含むガ/ ヲ/ニ格以外の項のタグ付与の問題がある．これについて検討するために，NAIST テキストコーパスの一部，136 記事を対象にガ/ヲ/ニ格以外の表層格（カラ/ヘ/ト/ヨリ/マデ/デ）のタグを試験的に付与し，どのような結果となるかを調べた．ただし，項の出現箇所に制限を加えずに作業を行った場合，作業者は各述語に対し文章全体を対象に格要素を探す必要があるため，すべての項に網羅的にタグ付与できるかどうかがわからない．そのため，今回は項を同定する範囲を述語と同一文内に限定し，その中で網羅的に項のタグ付与を行った．表 5 に作業者 2 人が付与したタグの個数を表層格ごとにまとめる．表 5 より，ガ/ヲ/ニ以外の項についてもある程度の個数が付与可能なように見えるが，このうち文 (9) や文 (10) のような，複数の述語が同一表現を項として持つ並列や文内ゼロ照応など，明示的にタグを付与すべき現象がどのくらい出現しているかを人手で調査したところ，作業者 1 と 2 でそれぞれ 16 回と 31 回であった．つまり，項のタグ付与の対象を同一文内に限定した場合，ほとんどの項は係り受け関係にあり，かつ明示的にタグ付与対象となる格助詞を伴い出現するため，今回人手でタグ付与作業を行った結果のほとんどは機械的に処理できる問題となる．

表 5: ガ/ヲ/ニ格以外のタグ付与結果

表層格	カラ	ヘ	ト	ヨリ	マデ	デ	計
作業者 1	133	9	260	17	32	374	825
作業者 2	130	11	311	17	22	405	896

(9) **台北**$_i$ では、屋外のスタジアムも満員に なり$_{デ:i}$、失神者が 出た$_{デ:i}$ ほど。

(10) ...「新民主連合」は六、九の両日に**総会**$_i$ を 開き$_{ヲ:i}$、離党問題などの対応を 話し合う$_{デ:i}$ ことにしており、党内調整は大きなヤマ場を迎える。

このため，文を越えて述語と任意格の関係を付与することを考慮する必要があるが，どのような基準でその作業に取り組めばよいのかは自明ではなく，今後さらに検討する必要があると考えている．

参考文献

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

[2] Hasida, K.: GDA 日本語アノテーションマニュアル 草稿 第 0.74 版 (2005). http://i-content.org/gda/tagman.html.

[3] Hirschman, L.: *MUC-7 coreference task definition*. Version 3.0 (1997).

[4] Jackendoff, R.: *Semantic Structures*, Current Studies in Linguistics 18, The MIT Press (1990).

[5] Kingsbury, P. and Palmer, M.: From TreeBank to PropBank, *Proceedings of the 3rd International Conference on Language Resources and Evaluation* (*LREC-2002*), pp. 1989–1993 (2002).

[6] Mitkov, R.(ed.): *Anaphora Resolution*, Studies in Language and Linguistics, Pearson Education (2002).

[7] Palmer, M., Gildea, D. and Kingsbury, P.: The Proposition Bank: An Annotated Corpus of Semantic Roles, *Computational Linguistics*, Vol. 31, No. 1, pp. 71–106 (2005).

[8] Poesio, M.: Discourse Annotation and Semantic Annotation in the GNOME Corpus, *Proceedings of the ACL 2004 Workshop on Discourse Annotation*, pp. 72–79 (2004).

[9] 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).

[10] 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).