

# 日本語述語項構造アノテーションに関わる諸問題の分析

松林 優一郎<sup>1,a)</sup> 飯田 龍<sup>2</sup> 笹野 遼平<sup>2</sup> 横野 光<sup>3</sup> 松吉 俊<sup>4</sup> 藤田 篤<sup>5</sup> 宮尾 祐介<sup>3</sup> 乾 健太郎<sup>1</sup>

概要：日本語の述語項構造アノテーションコーパスは、形態素解析や係り受け解析技術の場合と同様、述語項構造解析技術の発展に大きく貢献した。一方で、コーパスの作業員間一致率や定性的な分析からは、既存のアノテーションガイドラインに未だ改善の余地があることが分かる。本報告では、述語項構造アノテーションのガイドラインのさらなる改善を目指し、既存コーパスの仕様策定に関わった研究者や仕様の改善に関心のある研究者らの考察をもとに現状のガイドラインの問題点を洗い出し、整理・分析した結果を報告する。また、分析を通じてコーパスアノテーションガイドラインの作成・改善の方法論についても議論する。

## 1. はじめに

述語項構造は、文章内に存在する述語と、その述語が表現する概念の構成要素となる複数の項との間の構造である。例えば次の文、

(1) [太郎] は [手紙] を 書いた。

では、述語「書く」に対して、「太郎」と「手紙」がこの述語の項であるとされる。また、述語が表現する「書く」という概念の上で、それぞれの項の役割は区別される。役割を表すためのラベルは用途に応じて様々であるが、例えば、ここでの「太郎」には、「ガ格」「動作主」「書き手」などのラベル、「手紙」には「ヲ格」「主題」「書かれる物」などのラベルが与えられる。このように、述語に関わる構成要素を構造的に整理する事によって、複雑な文構造・文章構造を持った文章において「誰が、何を、どうした」のような文章理解にとって重要な情報を抽出することができる。このため、述語項構造の解析は、機械翻訳、情報抽出、言い換え、含意関係理解などの、複雑な文構造を取り扱う必要のある言語処理において有効に利用されている [14], [28]。

述語項構造解析においても、近年、形態素解析や構文解析などで行われている方法と同様に、人手で作成した正解解析例をもとに、統計的学習手法によって解析モデルを作成する方法が主流となっている [17]。述語項構造をタグ付けしたコーパスとしては、日本語を対象にしたも

のでは、京都大学テキストコーパス (KTC) [12] の一部に付けられた格情報 [11], [36] や、NAIST テキストコーパス (NTC) [8], [43], GDA コーパス [32], 解析済みプログコーパス (KNBC) [44], NTC の基準に従って BCCWJ コーパスに述語項構造情報を付与したデータ (BCCWJ-PAS) [33] などがあり、英語を対象にしたものでは、PropBank [24], FrameNet [26], NomBank [21], OntoNotes [7] などが主要なコーパスとして挙げられる。過去十年間の述語項構造解析技術の開発は、まさにこれらのデータによって支えられてきたとって過言ではない。

しかしながら、日本語の述語項構造コーパスは、その設計において未だ改善の余地を残す状況にあると言える。第一に、比較的高品質な述語項構造がアノテーションされた英語のコーパスに比べて、日本語を対象とした述語項構造のアノテーションは、省略や格交替、二重主語構文などの現象の取り扱いのほか、対象述語に対してタグ付けすべき項を列挙した格フレームと呼ばれる情報の不足などにより、作業員間のアノテーション作業の一致率に関して満足の得られる結果が得られていない。例えば、現在ほとんどの研究で開発・評価に利用されている NTC に関して、飯田らは、作業員間一致率や作業結果の定性的な分析を踏まえれば、アノテーションガイドラインに少なからず改善の余地があるとしている [43]。また、我々は、述語項構造アノテーションの経験のない日本語母語話者一名を新たに作業員とし、KTC, NTC のアノテーションガイドラインを熟読の上で新たな日本語記事に対して述語項構造アノテーションを行ったが、KTC, NTC のどちらのガイドラインにおいてもタグ付けを一意に決めることの出来ないケースが散見された。述語項構造のようにその他応用解析の基

<sup>1</sup> 東北大学  
Tohoku University, Sendai, Miyagi 980-8579, Japan  
<sup>2</sup> 東京工業大学 Tokyo Institute of Technology  
<sup>3</sup> 国立情報学研究所 National Institute of Informatics  
<sup>4</sup> 山梨大学 University of Yamanashi  
<sup>5</sup> 公立はこだて未来大学 Future University Hakodate  
a) y-matsu@ecei.tohoku.ac.jp

盤となる構造情報については、これに求められる一貫性の要求も高い。したがって、今後、述語項構造の分析や解析器の開発が高水準になるにつれて、既存のコーパスを対象とした学習・分析では十分な結果が得られなくなる可能性がある。そのような問題を防ぐためには、現状のアノテーションガイドラインにおいて判断の揺れとなる原因を洗い出し、継続的にコーパスの質を改善していく必要がある。

第二に、より質の高いアノテーションを目指してガイドラインを改善することを考えた場合、それぞれの基準をどういった観点で採用したかが明確に見てとれるような、論理的で一貫したガイドラインが必要となるが、KTC, NTCなどの既存のアノテーションガイドライン [37], [42] や関連論文 [8], [11], [36], [43] を参照しても、個々の判断基準の根拠が必ずしも明確には書かれていない。典型的に、アノテーションガイドラインの策定時に議論される内容はコーパス作成者の中で閉じた情報となることが多く、その方法論や根拠が明示的に示された論文は少ない。このため、付与すべき内容の詳細をどのように考えるかという、アノテーションそのものの研究が発展する機会が失われているという現状がある。また、KNBC や BCCWJ-PAS のように既存のガイドラインに追従して作られるコーパスの場合、新規ドメインに合わせるなど一部仕様が再考されるものの、アノテーションの研究は一度おおまかにその方向性が決まってしまうと、再考するための情報の不足もあり、本質的に考えなければならない点が据え置かれ、さらに詳細が議論されることは稀である<sup>\*1</sup>。

そこで、本研究では、この二つの問題を解消するために、既存のコーパスのガイドラインにおける相違点や曖昧性の残る部分を洗い出し、どのような部分に、どのような理由で基準を設けなければならないかを議論し、その着眼点を明示的に示すことを試みた。具体的には、(i) 既存の述語項構造コーパスの開発者、アノテーション作業員、また既存の仕様に問題意識を持つ研究者を集め、それぞれの研究者・作業員が経験的に理解している知見を集約したほか、(ii) 既存のガイドラインに従って新たな文章群へのアノテーションをあらためて行った結果に基づいて議論を行い、論点を整理した。(iii) これらをふまえ、述語項構造に関するアノテーションをどう発展させるべきか、どの点を吟味すべきかという各論とともに、アノテーション仕様を決める際の着眼点としてどのようなことを考えるべきかという議論も行った。本論文ではこれらの内容について、それぞれ報告する。

次節以降では、まず、2 節で述語項構造のアノテーションの先行研究を概観し、3 節で今回特に比較対象とした

NAIST テキストコーパスの述語項構造に関するアノテーションの仕様を紹介する。4 節で研究者・作業員が集まった際の人手分析の方針を紹介し、5 節で分析した事例を種類ごとに紹介する。さらに、6 節で述語項構造アノテーションを通じて考察した人手アノテーションに関する方法論について報告する。最後に 7 節でまとめと今後の課題を述べる。

## 2. 関連研究

述語項構造を解析したコーパスとしては、日本語文章に対するものに、京都大学テキストコーパス (KTC), NAIST テキストコーパス (NTC), GDA タグ付与コーパス (GDA), KTC 準拠のタグ付けをブログ記事に対して行った解析済みブログコーパス (KNBC), 日本語書き言葉均衡コーパス (BCCWJ) に対して NTC 準拠のタグ付けを行ったコーパス (BCCWJ-PAS) などがある。英語を対象としたコーパスとしては、FrameNet, PropBank, NomBank, OntoNotes などが主要なものとして挙げられる。特に、NTC, FrameNet, PropBank, NomBank などは、比較的多くの文章事例を含むことから、これまでに、様々な解析器の学習データとして用いられてきた [9], [17], [30], [31]。

表 1 に、各コーパスの特徴を示した。コーパス間の主な仕様の差としては、文書ドメイン、述語-項関係を表すラベル、格フレーム辞書の有無、文外の項に関する取り扱いの有無など挙げられる。

コーパスの文書ドメインは、従来、新聞記事を中心に整備されてきたが、係り受け解析等のその他の技術同様、教師あり学習によって開発された述語項構造解析器の精度が、学習データの文書ドメインに依存するという結果 [3] から、近年は複数文書ドメインへのアノテーションが進みつつある (BCCWJ-PAS, KNBC, OntoNotes など)。

述語-項関係ラベルとしては、文中の統語的なマーカを関係ラベルに利用した表層格、項のより意味的な側面を取り扱った意味役割ラベル等のバリエーションがある。日本語の述語項構造アノテーションの主要なコーパスである KTC と NTC では、日本語の統語上の格関係マーカである格助詞を関係ラベルとして利用している。KTC では、述語が現れた時、それが伴っている助動詞・補助動詞等を含めた形 (出現形) に対して一般的に付随する格助詞を用いて、項にタグを付与する。

- (2) a. [太郎<sub>ga</sub>] が [本<sub>wo</sub>] を 買う。  
b. [この本<sub>wo</sub>] は [太郎<sub>ni</sub>] に 買ってほしい。

上の例では、下線部が述語表現、[] 括弧で囲まれた部分が項、その内部の下付き文字が格関係ラベルを表す。以降、特に断りのない限りは、例文での項構造はこのように表す。一方で、NTC では、述語の原形に対して付随する格助詞

\*1 公開されているガイドラインを確認する限りでは、KNBC 作成時には格関係に関するガイドラインは再考されていない。BCCWJ-PAS の仕様は、機能語相当表現の判別に辞書を用いる点と、タグ付与の際に既存の格フレームを参照する点をのぞいて、NTC の仕様とおよそ同等である。

表 1 述語項構造コーパスの比較：OntoNotes 4 の名詞述語情報は、フレーム情報改善のため、一時的にデータから除外されている。BCCWJ-PAS は (小町, 飯田 2011) で報告されたデータ。2013 年 10 月現在, Yahoo!知恵袋コアデータ約 6,400 文のみ公開されている。

コーパス名	事例数	ドメイン	述語-項関係	格フレーム	名詞述語	文内の項	文外の項
KTC 4.0	5,000 文	新聞	表層格 (出現形)		✓	✓	✓
KNBC	4,186 文	ブログ	表層格 (出現形)		✓	✓	✓
NTC 1.5	約 40,000 文	新聞	表層格 (原形)		✓	✓	✓
BCCWJ-PAS	約 19,000 文	複数	表層格 (原形)		✓	✓	✓
GDA	約 37,000 文	新聞	主題役割				✓
FrameNet 1.5	173,018 事例	複数	意味役割	✓	✓	✓	
PropBank	約 113,000 事例	新聞	意味役割	✓		✓	
NomBank	114,576 事例	新聞	意味役割	✓	✓	✓	
OntoNotes 4	209,505 事例	複数	意味役割	✓	✓	✓	

を使ってタグ付与を行う。

- (3) a. [太郎<sub>ga</sub>] が [本<sub>wo</sub>] を 買う .  
b. [この本<sub>wo</sub>] は [太郎<sub>ga</sub>] に 買ってほしい .

この方法は、使役・受身・願望など、格の交替が起こる表現の間で格のラベルを正規化することで、表層格に、主題役割のようなより意味機能的な側面を持たせるを試みたものと捉えることができる。ただし、5.2.5 節でも述べる通り、この二つについては、相手側には含まれない情報をそれぞれ持っており、どちらの方式がより適切かはアプリケーションによっても異なるため、一概に優劣を決めることは出来ない。出現形アノテーションにおける格交替の情報を補う研究として、自動的に収集された出現形の格フレームの間で、格ラベルの交替がどのように起こるかを自動的に対応付ける研究 [27] がある。

英語に対する主要なコーパスでは、述語と項の間のより詳細な意味関係をとらえる、意味役割ラベルが用いられる。これは、例えば、Agent, Theme, Goal などの主題役割 (thematic roles) のように、項の述語横断的な意味機能を扱いたい場合に有用である。また、日本語でのアノテーションではあまり取り扱いのない、必須格と周辺格の区別についても扱っている。日本語のコーパスでは、唯一、GDA が主題役割を取り扱っているが、タグ付け対象が文外のゼロ照応関係にある項に絞られており、述語項構造に見られる現象を網羅しているとは言い難い。小規模な日本語文章への意味役割の試験的な付与例として林部ら [40] や松林ら [18] の研究が挙げられるが、林部らの研究では、作業員間一致率が F 値で 67%前後と低く、実用に至っていない。松林らの研究は、必須格に対する一致率は 91% と高いが\*2、タグ付けに必要なフレーム辞書のサイズが未だ小さく、実用には至らない現状である。

加えて、英語のコーパスでは、それぞれの述語が取り得る格を列挙した格フレーム辞書と呼ばれる資源を構築する

\*2 松林らの研究では、文外の項に対するアノテーションを行っていない点に注意されたい。

のが一般的な手法である。格フレーム辞書は、大規模な生コーパスの観察により、タグ付けに先立って構築される。アノテータは格フレーム辞書を参照しながら項構造のタグ付けを行うことにより、タグ付けの揺れを抑えることができるため、高い作業員間一致率を得ることができる。日本語の場合、英語に比べて項の省略が多く、また、英語のコーパスでは行っていない文をまたいだ項のアノテーションを行っているなど、アノテータが確認しなければならない領域が相対的に広いと、英語の場合と一致率の単純な比較は出来ないが、PropBank の項アノテーションに関する一致率は、周辺の格を含める場合で kappa 値で 0.91、含めない場合で 0.93 と極めて高い [24]。また、含意関係認識タスクのために FrameNet 準拠のコーパスアノテーションを行った研究では、意味役割のタグ付けに関する一致率が 91% であったとしている [2]。これに対して、明示的な格フレーム辞書を持たない NTC では、一致率が 83% 前後と相対的に低い。KTC では、ガイドラインを安定化させた段階での格関係アノテーションの作業員間一致率を 85% と報告している [36]。NTC の仕様準拠する形で BCCWJ に対するアノテーションを行った研究では、アノテータが既存の格フレーム辞書を参照しながら作業を行うことによって作業員間一致率に一定の改善を得ることが出来たとしている [33]。

英語を対象としたコーパスにおいては、一般に、文をまたいだ項についての取り扱いがない。これは、日本語が項の省略を頻繁に伴うのに対して、英語における項の省略が比較的少ないことに由来する。しかし、英語の文章においても、イベント間の照応関係や推論的解釈により、同一文中には現れないが暗黙的に定まっている項があると解釈される場合もあるため、近年は、この問題を解消するための試みも研究されている [13], [22], [29]。

### 3. NAIST テキストコーパス

我々は、可能な限り多くの現象を網羅した分析を行うという観点から、これまでに、最も多くの文数にアノテーショ

ンが行われてきた NTC の仕様をベースとし、適時 KTC との対比を行いながら議論を進める方針とした。本節では、NTC のアノテーションガイドラインについて、本論文の理解に必要な範囲の内容を簡単に説明する。また、3.2 節では、NTC の作業仲間一致率について、我々があらためて詳細に分析した結果を述べる。

一般に日本語述語項構造アノテーションを行うにあたって同時に含まれる照応・共参照情報については、それ自体が難解な問題を多く含んでおり、それらの考察・理論化は、一つの大きな研究テーマに相当するものである。このため、照応・共参照アノテーションに対する考察は今後の課題とし、議論の対象外とした。

### 3.1 アノテーションガイドライン

ここでは、NTC のガイドラインについて概要を説明する。より詳細な内容については、5 節での個別の議論の際に必要なに応じて付け加える。ガイドラインの全容については、Web サイト [42] を参照されたい。ただし、同 Web サイトの内容は、ガイドライン開発過程の情報が入り混じっており、必ずしも公開版データ<sup>\*3</sup>の作業時の規定を反映していない。文書化されたガイドラインと公開版のデータに相違が見られる点については、NTC の開発者に連絡を取り、実際の作業がどのようなものであったかを確認した。

NTC では、動詞、形容詞、名詞句+助動詞「だ」、節末の名詞句を述語とみなし、その項構造を、述語原形に対する表層格ラベルを用いてタグ付けする。また、イベント性名詞についても、述語同様の項構造を考え、タグ付けを行う。

(4) a. [太郎<sub>ga</sub>] が [花子<sub>ni</sub>] に [リンゴ<sub>wo</sub>] を あげ た。

b. [県<sub>ga</sub>] の 現在の 一般事務 [職<sub>wo</sub>] の 採用 は 日本国籍が 要件。 ( イベント性名詞 )

項は、ガ・ラ・二格に相当するもののみにタグを付与する。タグの領域は、IPADIC [35] で定められる形態素分割における一形態素とする。項が名詞句の場合は、最も後ろの形態素をタグの範囲とする。述語がサ変動詞の場合は「サ変名詞+する」の「する」に、名詞句+「だ」の場合は、名詞句の最も後ろの形態素に述語タグを割り当てる。

機能語相当表現については、述語とはみなさない。同様に、動名詞、形容詞の副詞的用法、固有表現内の述語も述語とみなさない<sup>\*4</sup> ( 下線部はタグ付けしない箇所 )。

(5) a. 彼の 話に よる と、 ( 機能語相当表現 )

<sup>\*3</sup> NTC 1.5 版をさす。

<sup>\*4</sup> NTC ガイドライン Web 版には「機能動詞についてもタグ付け対象とみなさない」とあるが、実際には機能動詞の認定が難しいとの判断から、通常の述語と同様のタグ付けが行われた。

b. 行政と業界の もたれあい の 構図。 ( 動名詞 )

c. 点の取り方を よく 知っている。 ( 副詞用法 )

d. 野鳥を 守る 会。 ( 固有表現 )

受身、使役などの場合は、述語原形の格でタグを付与する。但し、これらの格交替によって、原形の場合は取らなかった格が新たにガ・二格として増えている場合は、述語に付随する助動詞を仮想的な述語とみなし、そこに追加ガ/二格などの格を割り当てる<sup>\*5</sup>。

(6) a. [私<sub>extra-ga</sub>] は [父<sub>ga</sub>] に 死な れた。

b. [私<sub>extra-ga</sub>] は [彼<sub>ga</sub>] に [リンゴ<sub>wo</sub>] を 食べ させる。

項が省略されている場合は、文章中から対象の項を探し、タグ付けする。文章中に候補となる名詞句が存在しないが、特定の項が埋まっていると認識できる場合、外界照応として、「外界 (一人称)」「外界 (二人称)」「外界 (一般)」の三通りのタグを割り当てる。

(7) a. [牡蠣<sub>wo</sub>] を 食べる ため、[太郎<sub>ga</sub>] は 広島へ 行った。 ( 項の省略 )

b. [ $\phi_{exo1-ga}$ ] そろそろ 帰る うと 思う。 ( 外界照応 )

二重に主語を取る構文においては、「A は B が V」を「A の B が V」として置き換えることが可能な場合は「ノ格」で付与、それ以外の場合は、「八格」と「ガ格」を用いてタグを付与とする<sup>\*6</sup>。

(8) a. [広島<sub>no</sub>] は [牡蠣<sub>ga</sub>] が うまい。

b. [太郎<sub>ha</sub>] が [花子<sub>ga</sub>] が 好き だ。

c. [彼<sub>ha</sub>] が [英語<sub>ga</sub>] が 読める。 ( 可能動詞 )

### 3.2 作業仲間一致率

ガイドラインの分析に先立ち、我々は、飯田ら [43] が用いたものと同じのデータを用いて、NTC の作業仲間一致率を更に詳しく分析した。その結果を表 2 に示す。一致率は、二名の作業仲間が 30 記事にタグ付けした結果について、一名の結果を正解、もう一名の結果をシステムの推定と仮定した場合の適合率、再現率、F 値として算出した。このとき、推定されたトークンが、正解データにおいて項となる共参照クラスタの中のいずれかのトークンと一致すれば正解とした<sup>\*7</sup>。ただし、我々の評価方法では、飯田らの方

<sup>\*5</sup> 追加ガ/二格は公開版には含まれていない。

<sup>\*6</sup> 八格、ノ格は公開版には含まれていない。

<sup>\*7</sup> 各作業仲間がタグ付けした共参照クラスタが異なるため、表 2 は作業仲間二名のうちどちらを正解と見なすかによって僅かに結果が異なるが、どちら側からもおよそ同じような結果となったため、片側だけを記載した。

表 2 NAIST text corpus の作業員間一致率

種類	適合率 (%)		再現率 (%)		F1(%)
述語	92.34	(808/875)	94.61	(808/854)	93.46
イベント性名詞	96.48	(247/256)	79.17	(247/312)	86.97
ガ - 全て	80.88	(884/1093)	81.55	(884/1084)	81.21
ガ - 述語	82.35	(686/833)	82.95	(686/827)	82.65
ガ - 述語 - 係り受け関係あり	92.36	(435/471)	94.36	(435/461)	93.35
ガ - 述語 - 係り受け関係なし	69.34	(251/362)	68.58	(251/366)	68.96
ガ - イベント性名詞	76.15	(198/260)	77.04	(198/257)	76.60
ガ - イベント性名詞 - 係り受け関係あり	83.33	(55/66)	83.33	(55/66)	83.33
ガ - イベント性名詞 - 係り受け関係なし	73.71	(143/194)	74.87	(143/191)	74.29
ヲ - 全て	88.32	(416/471)	93.48	(416/445)	90.83
ヲ - 述語	89.92	(330/367)	95.38	(330/346)	92.57
ヲ - 述語 - 係り受け関係あり	96.25	(308/320)	98.40	(308/313)	97.31
ヲ - 述語 - 係り受け関係なし	46.81	(22/47)	66.67	(22/33)	55.00
ヲ - イベント性名詞	82.69	(86/104)	86.87	(86/99)	84.73
ヲ - イベント性名詞 - 係り受け関係あり	74.07	(20/27)	95.24	(20/21)	83.33
ヲ - イベント性名詞 - 係り受け関係なし	85.71	(66/77)	84.62	(66/78)	85.16
ニ - 全て	68.71	(112/163)	86.15	(112/130)	76.45
ニ - 述語	72.41	(105/145)	88.98	(105/118)	79.85
ニ - 述語 - 係り受け関係あり	82.61	(95/115)	89.62	(95/106)	85.97
ニ - 述語 - 係り受け関係なし	33.33	(10/30)	83.33	(10/12)	47.62
ニ - イベント性名詞	38.89	(7/18)	58.33	(7/12)	46.67
ニ - イベント性名詞 - 係り受け関係あり	16.67	(1/6)	100.00	(1/1)	28.57
ニ - イベント性名詞 - 係り受け関係なし	50.00	(6/12)	54.55	(6/11)	52.17

法と異なり、述語やイベント性名詞の位置が不一致の場合は、それらにタグ付けされた全ての項を不正解とした。

分析は、格ごとに、係り受け関係の有無、述語・イベント性名詞の別に分けて行った。結果、格ごと、または省略の有無によって、一致率にかなりのばらつきがあることが分かった。特に、項が省略される現象であるゼロ照応を伴う事例では、格の種類横断的に一致率が低い。顕著に低い値を示すのは、ゼロ照応のヲ格・二格、および、イベント性名詞に関する二格であるが、これらは事例数自体が少ないため、この結果がガイドラインの不備によるものかどうかを確かめるには、あらためて事例を収集し検証する必要がある。

#### 4. 論点の収集方法

本節では、既存コーパスのガイドラインにおける問題点を洗い出すために、我々が取った方法を説明する。ガイドラインの問題点を収集するための具体的な方法論は確立されていないため、今回は、(i) 既存のガイドラインを利用して新規アノテーションを行い、曖昧な箇所を探るという方法と、(ii) NTC・KTCの仕様策定、NTC、KTCを用いた応用処理に関わった研究者、述語項構造アノテーションの仕様に対して問題意識を持つ研究者が経験的に持つ知見を集約するという方法の、二つの方法を取った。

前述のとおり、本論文で取りまとめる考察は、NTCのアノテーションガイドラインを基準に行う。ただし、議論

上関連のある項目については、KTCのガイドラインとの対比を取り、より広範囲に考察を加えられるよう努めた。また、NTCやKTCのガイドラインにおいては、タグ付けする文書ドメインが限定されていることにより認知されなかった問題がある可能性も否定出来ないため、今回の論点収集の過程では、新聞ドメイン外の文に新たにアノテーションをすることを試みた。議論の対象となる題材は、述語項構造アノテーションの経験がない一般人の日本語母語話者1名、NTC・KTCの仕様策定関係者3名、述語項構造アノテーションの仕様に対して問題意識を持つ言語処理研究者6名の計9名によって、具体的に以下の手順で収集した。

- (1) 述語項構造アノテーションの経験がない日本語母語話者1名を新規アノテーションの作業員とする。作業員には、NTCのアノテーションガイドラインを熟読してもらい、その後、基本的なタグ付け方法について指導を行う。
- (2) Wikipedia, BCCWJよりサンプリングした例文に対して、NTCのガイドラインに従い、作業員が述語項構造をタグ付けする。判断に迷いが出た事例は、取りまとめて著者らに報告する。
- (3) 報告された事例について、著者らが、NTC・KTCのガイドライン、及びNTCデータ内の実際のタグ付け例と照らし、簡潔に解決可能かどうか確かめる。解決可能な場合、ガイドラインを更新し、解決案の説明と、

具体例を加える。解決不可能なものは、議論対象の分類表に加える。このとき、NTC と KTC の間での取り決めの対比も行う。

- (4) 作業者は、新しいガイドラインと未解決問題の分類表を持ち、作業済みのデータを修正する。1,000 文程度になるまで、新しい文章セットを受け取り (2) に戻る。
- (5) NTC・KTC の仕様策定に関わった研究者、既存の仕様に問題意識を持つ研究者ら計 8 名（著者らを含む）の意見を参考に、研究者が経験的に理解している仕様上の改善点を、(1)～(4) の工程で出来た議論対象の分類表に追加する。また、新たに用意した BCCWJ 上の記事 20 記事程度<sup>\*8</sup> に対して、上記 (1)～(4) の工程で改善したガイドラインを見ながら実際にアノテーションを行ってもらい、問題となった点を、議論対象の分類表に加える。

以上の方法で収集・整理した 4 種 15 項目の論点（5 節、表 3 を参照）について、著者らが議論を交わし、結果として得られた知見をまとめ上げた。

## 5. 個別の論点

本節には、4 節の方法によって収集されたガイドライン策定上の論点に関して、研究者間で議論した結果をまとめる。まず、我々は、収集された問題をおおまかな種類ごとに分別し、結果、4 種 15 項目の論点を得た。表 3 にその一覧を示す。内容としては、述語の認定基準、格の取り扱い、格や格フレームの曖昧性の問題といった、既存のコーパスに本質的に潜んでいた問題のほか、新聞ドメイン以外で新たに見られた現象もある。以下では、それぞれの論点について、議論の詳細を記す。

各論点に対する議論は、議論に参加した研究者が種々のアノテーションタスクの設計を通して知る経験的な知見にもとづいて行われる。我々の目的の一つは、これら設計時の基本的な理念とガイドライン上の取り決めの対応関係を集約することであるので、議論の過程で現れたガイドライン策定上の基本原則については、6 節にあらためて取りまとめる。

### 5.1 タグ付けすべき述語の認定基準

#### 5.1.1 述語項構造を重要視すべき述語とそうでない述語

文章中の述語は、その全ての述語項構造が等しく重要性を持つわけではなく、一部の述語に関しては、その述語項構造を解析する重要性が低いものもある。例えば、以下

<sup>\*8</sup> コアデータ内の、書籍、雑誌、白書、Yahoo!知恵袋、Yahoo!ブログをドメインとする記事の冒頭 10 文程度を利用した。OC01.00006, OC01.00472, OC01.00485, OC01.01765, OC01.02071, OW6X.00007, OW6X.00009, OW6X.00016, OY03.04233, OY03.04343, OY04.01354, OY14.02901, PB13.00021, PB14.00016, PB14.00057, PB19.00011, PM11.00031, PM11.00207, PM11.00223, PM11.00226

の文、

- (9) 驚いてはいられない。

において、「驚く」は文の内容上その項構造の解析が重要になるが、一方の「いる」のほうは、より機能的な述語であり、項構造を捉えるというよりは、むしろ「てはいられない」という 1 フレーズを機能的な表現とみなす方が自然と考えられる。述語項構造そのものを解析する重要度の低い述語に関しては、アノテーションコストの観点からも、解析器の評価をより重要度の高い項構造だけで適切に行えるようにするという点からも、区別して取り扱いたい。

述語項構造の重要度に関する問題として、本論文では、

- (a) 複合語  
(b) 機能語相当表現  
(c) 機能動詞構文・格交替を伴う機能表現

を取り上げる。これらは、文章中にありふれた事象のため、アノテーションコストに対する影響も大きい。以下では、それぞれの項目について、どのように取り扱うべきかについての議論の結果をまとめる。

- (a) 複合語：

以下のように、述語となりうる語の後ろに項が追従する形からなる複合語を考える。この場合、項自体がその複合語の主辞であるため、これら語の内部に現れる述語と項の意味関係は、そのまま項の意味を修飾する構造となっている。この形では、一般に、項の部分単体での語の意味はそれほど重要ではなく、複合語全体のかたまりの意味となって初めて実用的な意味を持つ場合が多く、内部構造を分解して解析することの重要度はその他の項構造と比べて低そうに見える。

- (10) a. 作業 [者 *ga*]  
b. 書き [手 *ga*]  
c. 輸入 [品 *wo*]  
d. 提案 [手法 *wo*]

NTC や KTC では、これらの複合語に関しては、全て内部の項構造をタグ付けしているが、このような表現は出現頻度も高く、アノテーションコストに対して占める割合も高い。従って、もし、応用処理の観点から見て重要度の低い関係とするならば、実際にこのような情報が必要なアプリケーションからのニーズを待って、後発的にタグ付けを始めるのでも良い。

一方で、次の例文のように、述語部分が主辞となる場合や、二つ以上の項を伴う複合語、複合語の外側にも項を取る場合などは、一般に項が内容語となるため、分解して項構造を考えることに通常と同様の価値があるとも取れる。

- (11) a. [計算機 *wo*] 使用

表 3 述語項構造アノテーションのガイドライン設計に関わる論点

カテゴリ	論点
タグ付けすべき述語の認定基準	述語項構造を重要視すべき述語とそうでない述語
	名詞のイベント性認定
	述語が複合語である場合の分解
格の取り扱い	二格の「必須格」性
	可能形・二重格構文・持主受身
	使役・受身・ムード・授受表現・機能動詞で追加される格
	慣用表現
	格交替と表層格タグの種類 (KTC 方式と NTC 方式)
	項としての形容詞 (二格相当)
格及び格フレームの曖昧性解消・必須項の見落とし	A の B, 連体節, ゼロ照応等における格フレームの曖昧性
	格フレーム辞書とアノテーションの一貫性
	非文へのアノテーション
新聞ドメイン以外で見られた現象	述語の省略
	疑問文の照応
	音象徴語

b. [計算機 <sub>wo</sub>] 使用 [者 <sub>ga</sub>]

表現)

c. [計算機 <sub>wo</sub>] の 使用 [者 <sub>ga</sub>]

f. すぐ に 食べ なければ なら ない。(モダリティ表現)

g. ジム に 通う よう に な った。(モダリティ表現)

ただ、接尾辞などのひときわ判断が容易なものを除いては、どの複合語が内部の項構造の価値が薄いかについて判断することは容易ではないため、個別に判断することは避けたい。例えば、その代わりに、作業コストを下げ一貫性を保つための工夫として、複合語内部の項構造関係はほとんどの場合に一意に定まる事に着目し、複合語内部の述語-項関係については、辞書的にその構造を管理しておくことなどが考えられる。そのようにすることで、文章中の事例ごとにタグ付けを行う必要がなく、作業コストが低下する上に、アノテーションの一貫性も保たれる。仮にこの方法をとった場合、複合語の外側に項が出現する際には、複合語内部の述語を対象に外側の項のみアノテータが格関係を付ければよいことになる。

(b) 機能語相当表現 (モダリティ等) :

次の例文の下線部の述語は、助詞相当表現やモダリティ表現の一部と考えるのが自然である。

(12) a. 彼 の 話 に よると, その 店 は とても 有名 らしい。(格助詞相当表現)

b. 夏休みの 課題 で 蝉 に ついて 調べ た。(格助詞相当表現)

c. 気温 が 上がる に したがっ て, だん だん と 汗 が で て き た。(接続助詞相当表現)

d. 見つけた と いっ て も, これは かなり 小さい もの です。(接続助詞相当表現)

e. 驚い て は い られ ない。(モダリティ

これについて、NTC では、例えば「通うようになる」の「なる」に対して、「機能語相当」のタグを付けることで区別している。ただし、網羅性を保証できないとの観点から、配布版 (1.5 版時点) にはこの情報は含まれていない。一方、KTC では、このような機能語相当の動詞に関しても普通の述語と特に区別なく述語タグが付与されている。

助詞相当表現やモダリティ表現は、内容語の慣用表現 (5.2.4 節) と同様に、句として強く結びつくことで、非構成的な意味を形成している。たとえば、(12a) に見られる「によると」は、このひとかたまりで、情報の出所や判断の拠り所を表現する機能を持つ [23]。「によると」は文において 1 つの格助詞のように振る舞うので、この中の「よる」のガ格が何であるのかを考えるのは不自然である。

上の例文からは、それぞれ、下線部の述語を含む次のような機能表現を抽出することができる。

によると、について、にしたがって、といっても、ては いられない、なければ ならない、ようになる

機能表現を例外扱いするにあたり問題となるのは、どのような基準で機能表現とそうでないものを弁別するかということであるが、これらの機能表現は、言語学や言語教育の分野で研究されており、[23] や [5] などの辞書が出版されている。自然言語処理の分野で電子的に利用可能な辞書として、松吉らが編纂した機能表現辞書 [20] などが存在する。アノテーション作業前に、これらの辞書を用いてあらかじめ

め機能表現に印を付け、ほぼ自動的<sup>\*9</sup>に「タグ付けすべきでない述語」と認定することにより、作業コストを下げるができる。辞書には載っていないが、機能表現と考えるべき表現を見つけた場合、作業時にその表現を辞書に追加するなど、既存の機能表現リストから漏れている表現を拡充することも必要であると考え。

(c) 機能動詞構文・格交替を伴う機能表現：

次の例文に見られるような機能動詞構文(13a)や授受表現(13b)における下線部bの述語は、直前の述語aに対して、アスペクトや態、ムード等の意味を付加する機能的な働きをするものと考えられている[19], [34]。

- (13) a. 事件 が 社会 に 混乱<sub>a</sub> を 与える<sub>b</sub>  
b. 私 が 彼 に サイン を 書き<sub>a</sub> て もらう<sub>b</sub>

このような述語に対して、下線部aとbの双方の述語項構造を付与することは、構造の重複となり、作業の価値が低い。また、述語bに関しては、機能的な振り舞いをするものであるから、述語項構造として取り扱う必要性も低い。したがって、より内容的意味を持つ述語aの方を基準の構造とし、bで追加される意味情報を態・アスペクト・ムードのマーカと解釈する方法も考えられる。これに関し、既存のコーパスのガイドラインは、NTCでは、機能動詞については、通常の述語と同様にタグを付与し、「もらう」などの表現には述語タグをアノテーションしない、としている。KTCでは、機能動詞についてはNTCと同様に扱われ、「もらう」「ほしい」などの表現は述語の一部としてアノテーションされる(「サインを 書いてもらう」など)。

機能動詞や授受表現を特別に扱う際の問題点は、やはり、その表現と取り扱いの方法が網羅的に列挙できるか、という点にある。機能動詞に関するリストとしては、[39]などがあるが、現象を網羅するわけではない。従って、具体的な作業方法の一案としては、上記のようなリストを出発点として、予め、あるいは作業時に段階的に機能動詞・授受動詞等に関する述語のリストを作っていく、コーパス中の事例を自動チェックするような仕組みを用いることで、作業を簡素化・半自動化する方法が考えられる。

ただし、(13)の例でも見られる通り、これらの表現が使役・受身相当の機能表現の場合は、述語aが本来持つ格に加えて、使役格などの新たな格が追加される場合もある。この場合の取り扱いについては、5.2.3節と同様の議論となる。

### 5.1.2 名詞のイベント性認定

名詞に対して述語と同様の項構造をアノテーションする

<sup>\*9</sup> 一部の機能表現に対しては、機能表現かどうかの曖昧性を解消する必要がある。例えば、「コンビニによると、ついお菓子をたくさん買ってしまふ。」の「よる」は、述語項構造解析の対象とすべき述語である。

ことを考える場合には、その名詞が実際に何かしらのイベントを表しているかどうか問題となる。例えば、次のフレーズにおける、「施設」という語について考えてみる。

#### (14) a. 研究 施設

この、「施設」という語はサ変名詞であり、「施設する」という動詞が作れるが、ここで「研究施設」は施設した結果物であり、イベントではない。このような語にも、便宜的に述語項構造を割り当てることはできるが、文脈上イベントとして解釈できない語に関して、その項構造を解析することは、本質的にあまり意味が無い。むしろ、イベントとして解釈される「施設」と、そうでない「施設」を区別することのほうが、解析上はより意味のあることである。

NTCでは、名詞述語に関して、そのイベント性を区別するためのタグ(結果物/内容, もの, 役割, ズレ)を用意しているが、このタグ付けに関する明瞭なガイドラインは存在せず、内省に頼っているのが実情である。ここでの論点は、どのような基準を設ければ、これを内省に頼る以外の方法で判別できるかということである。あるいは、明確な基準を設けることが不可能であっても、閉じたデータ内においては一貫性を保つような方法を模索する必要がある。

我々の議論の中では、複合語と同じように、このような語が出現する度にチェックリストに追加しておき、アノテーション時に自動的に注意をうながすようなマーキングを行うことで一貫性を高めるという方法が挙がった。

#### 5.1.3 述語が複合語である場合の分解

NTCでは、述語は基本的に一形態素に対してタグ付けするとしているが、形態素の分割基準は、既存の形態素辞書を拠り所にするため、どのような辞書を使うかによって、述語単位の取り扱いが大きく異なってくる。表4には、いくつかの複合語について、IPADIC, JUMAN, UniDic辞書に基づく形態素分割の差を示したが、辞書によって、あるいは、単語によって、分割の位置は異なる。

このような語の扱いに関しては、次の二点が問題となる。

- (1) どのような形態素分割基準を基準とするのが、述語項構造を考える上で最も適切か、
- (2) ある形態素分割基準に基づいて複合語が二形態素以上に分割されたとき、複合語内部の述語はその全てがタグ付け対象として適切かである。しかし、どちらの問題も、現状で合理的結論を出すことは簡単ではない上、後に6節にも述べる通り、言語処理アプリケーションによっては、どの単位を述語として扱うのがよいか、また、どの程度複合語内部の項構造が必要となるかに異なりがある。例えば、含意関係認識タスクにおいては、表4の「立ち読み」や「消し 忘れ」がどのような理論に基づいて分割されているかにかかわらず、「私が、立って、本を、読む」ことや、「私が、ライトを、消そうとして、消すのを、忘れる」ことを理解する必要がある。

したがって、現状で完全な解決策を提示することは難し



表 4 IPA 辞書, JUMAN 辞書, UniDic による形態素分割の違い. 同じ品詞構成であれば, 同じ基準というわけでもない.

品詞構成	対象語	IPA	JUMAN	UniDic
VN+VN	立ち読み(する)	立ち読み	立ち読み	立ち読み
V(N)+V(N)	消し忘れ(る)	消し 忘れ(る)	消し 忘れ(る)	消し 忘れ(る)
V+V	歩き回る	歩き回る	歩き 回る	歩き回る
V+V	振り返る	振り返る	振り返る	振り返る
V+V	食べ歩く	食べ 歩く	食べ歩く	食べ歩く
N+VN	歯磨き(する)	歯磨き	歯磨き	歯磨き
N+VN	子供扱い(する)	子供 扱い	子供 扱い	子供 扱い
N+VN	二人乗り(する)	二人 乗り	二人 乗り	二人乗り
Adv+VN	ポイ捨て(する)	ポイ捨て	ポイ 捨て	ポイ捨て
Adj+N	悪影響(*する)	悪影響	悪 影響	悪 影響

いが, 当面の対処案として, まずは, ある特定の形態素分割辞書を使って分割された用言や体言のうち, イベント性があると判断するものは全てタグ付けするという方法を取り, 加えて, 5.1 節の複合語の項目で述べたような複合語内部の項構造を辞書的に管理する方法を, 必要に応じて一形態素に対しても適用することで, どのような形態素分割基準を用いた場合でも想定するアプリケーションの要求に対応できる柔軟な構造を取るという方針が挙げられる.

## 5.2 格の取り扱い

### 5.2.1 二格の「必須格」性

述語のそれぞれの項を, 主題役割のような意味役割のレベルで考えると, 「が」「を」に比べて, 助詞「に」を伴って出現する述語-項の関係には様々なものがある [38]. このうち, 初期段階の述語項構造アノテーションとして, 特別重要度が高いのは, タグ付け対象の述語そのものの概念を説明するために必須となる項目(必須格)である. 一般に, 助詞「に」を伴って出現する述語の項のうち, 必須の二格とみなされるのは, 動作による移動の着点や, 結果状態を表すものなどである. 一方, 時間, 様態などを表す「に」は, 述語横断的に利用可能な付加的修飾要素であるため, 周辺格などと呼ばれる.

しかし, 「が」「を」に比べて, 二格では, 必須格性の判断が容易ではないケースも多く存在する. 本論文では, 特に

(a) 必須格と周辺格の境界

(b) 二格の任意性

の二つについて取り上げる.

(a) 必須格と周辺格の境界: 例えば, 次の例

(15) a. 二つに 割る

b. こなごなに 割る

c. めちゃくちゃに 割る

を見ると, (15a) では, 二格は, 動作の結果状態を表しているように見えるが, (15b) や, (15c) のような表現になると, それが結果状態を指すのか, 動作(あるいは変化)の様態

を指すのかは極めて曖昧になり, 判断が難しくなる. 必須格と周辺格の区別については, 明確な基準を持って分けられる事例もあれば, 上記のように, どちらに属するとも言えない, ぼんやりとした事例も存在する. アノテーションを行う際に本質的に問題にしなければならないことは (i) 理論上どのようにタグ付けするのが合理的か, ということと, (ii) 揺れなく, 明確にタグ付けや評価が行える基準を設けなければならない, ということである. (i) の観点から言えば, もし上記のような必須格と周辺格の間の境界が本質的に曖昧なのであれば, 曖昧な状態を取り扱うことのできる表現にしておけば良い. 一方で, タグ付けや評価を行う場合は, 不確かなものは問題となる. 少なくとも, どの事例に関しては明確に区別可能であり, どの事例が本質的な曖昧さを含むのかを明らかにしておかなければ, 作業員間一致率や解析システムの評価時に, アノテーションやシステムの誤りであるのか, 本質的な曖昧性のために揺れているのかを区別できない.

この問題を解消するための方法として, アノテータが迷った事例に対しては, 迷ったことを示すマーカーを用意し, 対立候補と共にチェックをしてもらうことで, 明確な事例と曖昧な事例を区別しておく方法が考えられる. そうすることで, 評価用データとして用いる際もより厳密な評価を行うことができるようになる.

(b) 二格の任意性: 第二に, 文章中に存在しない二格を補う場合の問題がある. ある格が必須格だと判断した場合, それはすなわち, 仮にその格を埋める項が文章中に存在しない場合でも, 概念上は項が存在しているとみなすということである. しかし, 必須格と周辺格を一般によく知られている意味機能的な役割で分類しようとする, 動作の結果状態のように, 一般的には周辺格ではないと認識されている役割であっても, 述語によっては, 項が埋められている必要がある(暗に省略されている)と感じにくいケースもある.

(16) a. 信号が (φ二) 変わったので, 停車した.

b. 花瓶を (φ二?) 割った.

c. ボールが (φニ?) 落下する .

例えば, (16a) では, 信号が変わった結果の状態について, 文脈から何かしら明確な項を仮定する (赤に変わった, と仮定する) のが普通であるが, (16b) については, 特定の具体的な結果が指定されていなくとも, 「割る」の一般的な結果状態は「割る」という語の語義の中に初めから含まれているため, 意味は解釈できる. (16c) の「落下する」という動詞では, 二格で移動の着点を指定することはできるが, 必ずしも落下の結果どこかに到達している必要はないので, 二格が必須の項であるとは言いがたい. このような二格の任意性は, 述語ごとにそれぞれ判断が必要である. どのような基準で, 二格の任意性を認めるかについては, 現状では明確な基準は用意されていない.

また, 仮に, ある述語について, 二格の任意性が判定できたとしても, 実際の文中の事例について, 任意である二格が明示的に格助詞「に」を伴って出現していなかった場合, それが, 未定義なのか, 概念上存在しているのか, あるいは, 同一記事中の別の箇所に出現しているかどうかの判断も, 困難を極める. 例えば, 次の文

- (17) 衛星は, 落下し 始めた. 2 時間後, 太平洋で発見された .

の「落下する」の二格は, 未定義なのか, 文章中に存在しない「地球」なのか, それとも「太平洋」なのかは, 文脈をどのように解釈するかに依存する. このように, 文脈や事前知識に深く依存する問題については, 述語項構造アノテーションの範疇外としておき, それ以降の, 例えば推論モデル等で取り扱う問題と規定する考え方もありうる. 仮にそうした場合は, 明示的に格助詞と共に表れる場合や, 文脈上自明な場合を除いては未定義とすることになるだろう.

### 5.2.2 可能形・願望・二重ガ格構文・持主受身

可能動詞や可能形, 願望, 及び, いわゆる二重ガ格構文においては, 異なる意味機能を持った二つの格助詞「が」を伴うことがある.

- (18) a. 太郎は (が) 英語が/を 読める. (可能動詞)  
b. 太郎は (が) ブロッコリーが/を 食べられない. (可能形)  
c. 太郎は (が) ビールが/を 飲みたい. (願望)  
d. 太郎は (が) 足が 長い. (二重ガ格構文)

これについて, NTC では, 可能形の場合は原形に戻してタグを付与し, 「A は B が V」を「A の B が V」として置き換えることが可能な場合は「ノ格」で付与, それ以外の場合は, 「ハ」と「ガ格」を用いてタグを付与とされている.

- (19) a. [太郎 <sub>ha</sub>] は [英語 <sub>ga</sub>] が/を 読める.

- b. [太郎 <sub>ga</sub>] は [ブロッコリー <sub>wo</sub>] が/を 食べられない.

- c. [太郎 <sub>ha</sub>] は [ビール <sub>ga</sub>] が/を 飲みたい.

- d. [太郎 <sub>no</sub>] は [足 <sub>ga</sub>] が 長い.

しかし, そのような方法を取る場合, 次のような, ガ格・ヲ格の選択肢の範囲を限定する「は」の用法が現れたときに, タグ付けの対象を混同してしまい, 場合によっては二重の「ハ」となってしまう.

- (20) a. [ワイン <sub>ha?</sub>] は [ロゼ <sub>ga</sub>] が 美味しい.  
b. [私 <sub>ha</sub>] は [ワイン <sub>ha?</sub>] は [ロゼ <sub>ga</sub>] が 好きだ.  
c. [本 <sub>ha?</sub>] は 英語の [もの <sub>wo</sub>] を 読む.  
d. [私 <sub>ga</sub>] は [本 <sub>ha?</sub>] は 英語の [もの <sub>wo</sub>] を 読む.

上記のような例を考えると, 項の選択範囲を限定する「は」は述語横断的に利用できる周辺的な格と類推できるため, 必須格と周辺格を付け分ける現行のガイドライン上では, (19) におけるハ格と, (20a) におけるハの用法は明確に区別したい.

経験的に, 格のラベルと文中の実際の助詞が見た目上一致すると, アノテータはこうした混同を起しやすしい. したがって, これを避けるために「ハ」ラベルの名称を二つに分けるという方法が有効な可能性がある. ここでは, 例えば便宜的に (19) のハの場合を「属性所有のガ」, (20a)(20c) の場合を「限定ハ」と決めるような方法である. ラベルの名称を機能によって細分化するという方法は, 格助詞を直接格関係のラベルに用いる日本語の述語項構造アノテーションにおいては, 同じ助詞によって表される必須格と周辺格を区別する際に有効な手段であると考えられる.

一方, KTC の場合, 動作主体や経験者といった意味役割的な観念を用いて, 『二重のガとなるもののうち, 「は」「が」が動作主体や経験者である場合は, 用言からみて遠い方のガ格をガ2格とする』とすることで, 必須格と周辺格の混同を避けている. また, NTC でノ格に対応する「太郎は足が長い」などの表現は, 「は」を「が」に言い換えると不自然だとして, ガ・ヲ・ニなどの格助詞では言い表せない「外の関係」として定義している.

- (21) a. [太郎 <sub>ga2</sub>] は [英語 <sub>ga</sub>] が 読める.  
b. [太郎 <sub>ga2</sub>] は [ブロッコリー <sub>ga</sub>] が 食べられない.  
c. [太郎 <sub>ga2</sub>] は [ビール <sub>ga</sub>] が 飲みたい.  
d. [太郎 <sub>out</sub>] は [足 <sub>ga</sub>] が 長い.

これとは別に, 可能動詞には, NTC・KTC 双方におい

て格ラベルの組み合わせに曖昧性が出るという問題がある  
( (22) の a と b, c と d がそれぞれ曖昧)\*<sup>10</sup> .

- (22) a. [太郎 *ha*] は [英語 *ga/wo*] が/を 読める .  
(NTC 方式)
- b. [太郎 *ga*] は [英語 *wo*] が/を 読める . (NTC 方式)
- c. [太郎 *ga2*] は [英語 *ga/wo*] が/を 読める . (KTC 方式)
- d. [太郎 *ga*] が [英語 *wo*] を 読める . (KTC 方式)

この問題については、5.3.1 節で詳しく議論する。

### 5.2.3 使役・受身・ムード・授受表現・機能動詞で追加される格

NTC は、述語と項の間の格関係を、述語原形に対する表層格によって記述する。このような方法を取る場合、述語が使役・受身などの形を取った場合に、原形では対応のない格が出現する問題があり、これに対処する必要がある。

- (23) a. [私 *extra-ga*] が [太郎 *ga*] に 勉強  
さ せる .
- b. [彼 *extra-ga*] が [父 *ga*] に 死な れた . (迷惑受身)
- c. [私 *extra-ga*] が [彼 *ga*] に [ゲーム *wo*] を 壊さ れた . (持主受身)
- d. [両親 *extra-ga*] が [太郎 *ga*] に 勉強 し て ほし  
が つ い る . (願望)
- e. [私 *extra-ga*] が [彼 *ga*] に [本 *wo*] を 書い  
て もら う (授受表現)

この問題に関して、NTC では、上記のように助動詞や補助動詞を新たにマークし、追加ガ/ニ格を割り当てるとしている\*<sup>11</sup> . NTC のガイドラインでは、少数の助動詞・補助動詞に関して、具体的な事例を用いてアノテーション方法を指示しているが、これに加えて、機能動詞構文について 5.1 節で取り上げたような取り扱いをする場合は、先行する述語側で、追加の格関係を取り扱う必要がある。また、述語によっては、機能動詞によって格が追加されたと思なすべきか、受益格のような周辺の格と思なすべきか明確でないケースも存在する。

- (24) a. [事件 *extra-ga*] が [社会 *ga*] に 混乱  
を 与 え る (機能動詞構文)
- b. [彼 *extra-ni?/benefactory(peripheral)*] に [ジャム

\*<sup>10</sup> KTC のガイドラインでは、「基準として、可能形の動詞の対象(目的語)の格はヲ格、動作主体の格はガ格とするが、もっとも自然な格を選択する」としている。

\*<sup>11</sup> NTC 1.5 版時点で、公開版にはこのタグ情報は含まれていない。

*wo*] を取ってやる/あげる

特に、機能動詞や補助動詞については、表現の種類が多岐にわたるため、追加されている項が省略されている場合の見落としなどを抑制して作業の一貫性を高めるためには、これらの現象に関わる表現について、網羅的にかつ、統一的な扱いをする必要がある。これには、追加の格が存在する表現を一覧化し、チェックする仕組みを設けるのが好ましいと考えられる。

### 5.2.4 慣用表現

次の例のように、見た目上は、述語と格助詞を伴った項のようにも取れるが、実際には句として強く結びつくことで、一つの新たな意味を形成している慣用表現がある。

- (25) a. 私ガ/の気が滅入る
- b. 私のチームに手に入れたい
- c. 確認作業に骨を折る
- d. 彼の耳に入る

NTC では、どのような表現までが慣用表現と言えるのかの境界が厳密には規定できないのではないかという前提から、慣用表現かどうかを区別せずに、見た目上の述語に対してアノテーションをしている。KTC も同様に、慣用表現かどうかは区別せずにアノテーションを行っている。

これらの句に対して、述語項構造アノテーションのガイドラインが取り得る戦略としては、(i) NTC や KTC と同様に、慣用表現内部の述語項構造も全て分解してタグ付けする、もしくは(ii) 慣用表現は複数形態素にまたがる述語表現として特別扱いする、ということが考えられる。ただ、どちらの場合に関しても議論の余地がある。

(i) の場合は、まず、5.1.3 節の複合語の議論の時と同様、慣用表現内部の項構造は、出現事例ごとに異なるということはほとんどないため、同じ構造を何度もタグ付けする無駄が生じる可能性がある。また、慣用表現の表す意味は、比喩的な派生の結果、元の語句から構成的に組み上げられる意味と一致しないため、分解して項構造をタグ付けする意味が薄い。さらには、(25a)(25b) にも見られる通り、慣用表現によっては格の重複が起こり、どちらが、内容的に見て重要な格で、どちらが慣用表現内の「意味的重要度の低い」格かの区別が難しくなる。(25c) に見られるように、元々の述語(この場合、「折る」)に存在しなかった格(二格)が増える場合もあり、アノテーションに際して格フレーム辞書を用意した場合などには、分解された語のみの格フレームでタグを付与しようとするのが扱いが難解になる。

(ii) の場合は、ある句を、どのような基準で慣用表現とみなすかが問題となる。慣用表現を整理した既存の研究としては、佐藤の基本慣用語五種対照表 [41] や橋本らの OpenMWE: 日本語慣用語コーパス [6] などが挙げられる

が、佐藤の研究では、「慣用語の定義はいまだに決定的なものがない」としている。また、慣用表現全体を述語と見なすこととした場合には、(25d)のように、慣用表現内の項の一部を修飾する情報をどのように扱うかも問題となる。この例の「彼の」は、もし慣用表現を分解して考えた場合には、二格相当の句の一部となっているため、この関係にも何らかのラベルを用意するのが望ましいと考えられる。

この問題に対しては、まずは、コーパス内の慣用表現と思われる事例を集め、慣用表現を述語項構造という観点で見た場合にどのような現象が起こりうるのかを網羅的に収集する必要がある。そのため、まずは慣用表現内を分解した状態でタグ付けを行い、その上で慣用表現の取り扱いを決めるといった段階的なアノテーションが好ましい。また、実際に慣用表現をひとまとめにしたアノテーションを行う際は、機能語表現や機能動詞での議論と同様、慣用表現を辞書的に対象表現を管理するのが望ましい。

#### 5.2.5 格交替と表層格タグの種類 (KTC 方式と NTC 方式)

2 節で紹介したとおり、KTC は述語の出現形に対する格関係をタグ付けし、NTC は原形に対する格関係をタグ付ける。このため、格交替をともなって述語が出現する場合には、これら 2 つの基準では異なったアノテーションが行われる。出現形アノテーションと原形アノテーションは、互いに相手側には含まれない情報を持っており、どちらの方式がより適切かはアプリケーションによって異なる。

例えば、含意関係認識のような命題間の同一性を扱いたいタスクでは、(26) の a と b が同じ内容を表していることを捉えたい。このため、このような場合は、格交替を吸収する NTC 方式が有用である。

(26) a. [次郎 *extra-ga*] は [太郎 *ga*] に [ご飯 *wo*] を 食べられた。(NTC 方式)

b. [太郎 *ga*] が [ご飯 *wo*] を 食べた。(NTC 方式)

一方で、機械翻訳や文書要約などの表層的な形式をそのまま扱うことが可能なアプリケーションでは、受身や使役などが出現している場合、それをそのまま翻訳すれば良いため、必ずしも原形に戻す必要性はない。この場合、出現形に関する項に関して省略がある場合はそれを補完すればよい。そのような時にも述語を原形に戻そうとした場合、原形に対する格パターンを選択する際に処理を誤る可能性もあるため、無理に原形に戻す処理を行うことはリスクをともなう。このような場合には出現形でアノテーションを行う KTC 方式を採用するほうが望ましい。

(27) a. [太郎 *ni*] が来た。[りんご *wo*] を 食べられた。(KTC 方式)

b. [太郎 *ga*] が来た。[りんご *wo*] を 食べられた。

(NTC 方式: 受身のまま「太郎」を補う場合に、二格で補われるべきという情報を得られない)

これらに関連して、格交替前と格交替後の格の間の対応関係を取りたい場合に、KTC 方式の場合、この方式でアノテーションしたコーパスからは、この対応関係を直接学習出来ないため、対応関係を獲得するための新たな資源が必要となる。NTC 方式の場合、コーパス上にこの対応関係をタグ付けしていることになるので、見た目上そのような対応関係表は必要ないが、実際にはコーパス中に出現する交替現象は全述語の 1 割程度であるため、異なる格交替の振る舞いをするそれぞれの述語に対して、対応関係の学習に十分な量の交替事例が得られるとは限らない。出現形表層格における格交替関係については、10 億文規模の大規模なコーパスから自動獲得する方法も研究されているため [27]、格交替の扱いについては、今後どちらの方針でアノテーションすることが効果的かを検証する必要がある。

この検証を行うためのデータ作成の方法として、KTC 方式、NTC 方式の双方で同一文章にアノテーションを行う方法が考えられる。この場合のコストは、格交替が起こらない場合などの重複する作業は省略できるため、単純に倍というわけではない。ただし、効果的に対応関係を取るためのアノテーションの方法については今後検討する必要がある。

もう一つの方法は、仮にいくつかのデータが、アプリケーションによる要請などによって異なるタグセットを用いてアノテーションされたとしても、それぞれのスキーマによるアノテーションの結果を自然に統合し、互いにラベルセットを交換可能とする仕組みを考えることである。KTC と NTC の場合は、各述語に対する語義別の格フレーム辞書と、各語義に関する格交替の性質を網羅的に記述した辞書を用いて、この仕組が設計可能である。この方法を取れば、将来、主題役割などのラベルを導入する場合にも、既存のアノテーションの結果をマッピングすることで、最小限の再アノテーションによって、新たな結果を得ることが期待できる。ただし、このようなスキーマ間のラベルの対応を得るのは容易ではない。再アノテーションを避けるためには、異なるスキーマ間のラベルが事例ベースで一対一対応する必要があるが、各事例で適切な対応関係を得るためには、それぞれのスキーマが、お互いのラベルがエンコードしている情報の差を明確に意識し、その差が、追加情報によって将来的に埋められるよう綿密に設計されたスキーマでなければならない。また、格フレームや語義等も、共通の基盤データに基づいておく必要がある。さもなければ、それぞれのスキーマの理論上のずれや、格フレームのカバレッジ、語義の粒度のずれによる影響で、ラベル間の対応が、一対多、多対多の曖昧な関係となり、結局、再アノテーションを行わざるを得ないことになる。実際に、英

語圏では、異なる述語項構造コーパス間にタグ付けされた異なる情報を有用に利用しようと、資源間でのマッピングを試みた研究があるが、それぞれ異なる理念で設計されたコーパスであったため、格フレームやラベルの対応関係は多対多となり、再アノテーションを必要とした [15], [16] . したがって、仮に、アプリケーションからの要請や、段階的に情報を付加していく設計などによって、異なるアノテーションスキーマを使い分ける場合にも、将来の統合性をはっきりと意識した設計をしておくことが重要となる。例えば 5.2.1 節で述べたような、必須格と周辺格の区別などは、現状のガイドラインでは明確に取り扱われていないが、意味役割との親和性を考えれば、重要な事項である。

#### 5.2.6 項としての形容詞（二格相当）

次の二つの例文は、非常に似通った意味を表している。

(28) a. [服<sub>ga</sub>] が [赤<sub>ni</sub>] に 染まる .

b. [服<sub>ga</sub>] が 赤く 染まる .

どちらの文からも、我々は「服が赤くなった」という同一の結果状態を想像することができる。しかし、現状の表層格を用いたアノテーションでは、「赤く」という形容詞を用いた表現は項として認識されず、これら二文の間の項構造は異なるものになる。この違和感は、特に項の省略を伴う次のような例文に対して、どのような表現まで項として補うかという判断を行うときに、特に大きくなる。

(29) a. [真っ赤<sub>ni?</sub>] なペンキで、[服<sub>ga</sub>] が 染まっ てしまった。

b. [赤い?] ペンキで、[服<sub>ga</sub>] が 染まっ てしまった。

この問題は、我々が、表層格というラベルを用いて、述語とそれを取り巻く要素の間のどのような関係を取り扱おうとしているかを考える際の良い題材である。現状のスキーマでは、格助詞の表層的な違いとして認識できる粒度の意味関係しか取り扱っておらず、果たしてどのような意味機能をもったものならば、ガ格・ヲ格・ニ格との意味的対応関係が取れるのかについて、網羅的な結論を即座に出すことは難しい。しかし、もし、述語項構造を、述語と項の間の意味的関係の同一性を示すための表現として用いようとするならば、「名詞＋格助詞」や「形容詞」といった統語上の区分にかかわらず、同一の意味機能を持つものには同一の関係ラベルを与えるのがよいかもしれない。これは、将来発展的に、主題役割のような、より意味機能的なラベルを用いてアノテーションスキーマを設計しようとする際には、十分検討されるべき課題である。

### 5.3 格及び格フレームの曖昧性解消・必須項の見落とし

#### 5.3.1 A の B, 連体節, ゼロ照応等における格フレームの曖昧性

ある述語が複数の格フレーム候補を持つとき、その述語が、A の B・連体節・項のゼロ照応などの形を取った場合、アノテーション時にどの格フレームを選択すべきかについて曖昧性が生じる。

(30) a. 自他交替：パソコンの起動 → パソコンが起動する / パソコンを起動する

b. 道具格交替：ドアを開けた鍵 (誰かが) 鍵で開ける / 鍵が開ける

c. 他動詞 / 自動詞 + 使役：政府による経済再生  
[政府<sub>ga</sub>] が [経済<sub>wo</sub>] を再生する / [政府<sub>extra-ga</sub>] が [経済<sub>ga</sub>] を再生させる

d. その他：私が教える生徒 生徒を教える / (何かを) 生徒に教える

また、述語によっては、同一の意味機能を持つ項に対して、複数の格助詞が代替可能である場合がある。

(31) a. 私が / から 話す

b. 太郎に / から もらう

c. 風に / で 揺れる 花びら

d. 土台に / と くっつける

この例では、ガ・ヲ・ニの間で代替可能なものはないが、仮に今後タグ付け対象の格助詞を拡充することを考える際には、このような曖昧性を生み出す要素に対して、どのように一貫したアノテーションを行うかについて考慮する必要がある。

タグ付けに本質的な曖昧性が出る場合には、ある基準に基いて（例えば、出現頻度順や、アノテーションコストが低くなるように、などで）決めた規則に従って、タグ付けが一意に定まるようにするのが一般的である。しかし、前者の格フレームの曖昧性については、文脈によっては、どちらか一方の格フレームの方が他方での解釈よりもより自然な場合があり、そのような場合は、適切な解釈となる格フレームを選ぶのが好ましいと考えられる。一方で、文脈の曖昧さによっては、アノテータ間の意見が一致しない場合もありうるし、当然、本質的にどちらに解釈しても自然な場合もある。そのような事例に対しては、自動解析器の学習や評価時に適切な取り扱いができるように配慮しなければならない（どちらの解釈でも正解とするなど）。

このような場合を総合的に考えて対策を取るとすれば、例えば、(i) 事前に述語ごとの格フレーム辞書を用意しておくか、あるいは、代表的な格フレーム交替についての名称

を列挙しておき、(ii) アノテーション時に複数の格フレームで判断に迷うもの、あるいは本質的に曖昧なものについては、その交替の候補を列挙し、(iii) 自己の判断、もしくは規則に従った判断で選んだ格フレームで格関係をアノテーションする、という方法などが考えられる。

後者の、代替可能な格に関しては、今、仮に、ガ・ヲ・ニ・カラ・デ・トの6つの表層格を使ってアノテーションをしている場合を考えると、この6つのラベルに対して、ガ・ヲ・ニ > カラ > ト > デなどの半順序を与えることで規則的にこれを解消することもできるが、文中で、

(32) 彼には 私から 話しておく。

と格助詞「から」を伴って出てきている事例に対して、これを「ガ格」として正規化するためには、アノテータはタグ付け毎に、対象の述語に関する格フレーム辞書を想起し、格の交替関係を確認せねばならず、アノテーションのコストが大きい。したがって、少なくとも、述語が原形で使用されているもので項が格助詞を伴って現れる場合には出現形の格でタグ付けし、受身・使役などで格交替しているものや、ゼロ照応などで元の格助詞が不明なものに関しては、上記の半順序規則を適用するなどといった方法が好ましいと考えられる。ただし、タグ付け時に格を正規化しなかった場合に、曖昧な格のうちいずれの格でも本質的に正しいという取り扱いをするためには、別途格フレーム辞書等に述語毎の格の交替情報を記述しておくなどする必要があるのである。

### 5.3.2 格フレーム辞書とアノテーションの一貫性

NTC や KTC のアノテーションでは、被連体修飾詞やゼロ代名詞として出現する項など、明示的に格助詞を伴わなかったり、対象の述語と何らかの統語的關係を伴わない項に関してもタグ付けを行う。しかし、タグ付け作業時点では、ある述語の取り得る格(格フレーム)について参照できる辞書等が存在しなかったため<sup>\*12</sup>、アノテータは内省に頼りながら、文章中からその述語に足りない項を補う作業を必要とした。しかし、一つ一つの述語の格フレームの定義をアノテータの内省に頼る方法には限界があり、その影響は、3.2 節、表 2 の NTC の作業員間一致率においても、ゼロ照応項の不一致という形で顕著に見られる。

述語項構造アノテーションの一貫性を今以上に向上させるためには、予め各述語に対して正確な格フレーム辞書を定義しておくなどして、全てのアノテータが共有する共通の語彙知識ベースを整備する必要がある。実際に、NTC ガイドライン準拠のアノテーションを BCCWJ に対して行った研究 [33] では、既存の格フレーム辞書を一部参照することによって、作業員間一致率に一定の向上が得られた

<sup>\*12</sup> NTC では、NTT 語彙大系構文体系の辞書や竹内らの動詞語彙概念辞書などを参照して作業したものの、辞書に記述されている格パターンと新聞に出現する格パターンが必ずしも一致せず、逆に作業員が混乱したために、使用が中止された。

としている。

作業の上で参照する格フレーム情報は、できるだけ精緻なもの求められるが、一方で、大規模な文章に対する述語項構造アノテーションを行うにあたって、ある述語の様々な言語現象を網羅した、実用に耐えうる頑健な格フレーム辞書を人手で用意するには膨大なコストを必要とする。コストを抑えた方法として、大規模な文書データから、自動的に格フレームを獲得する研究が存在するが [10]、獲得したフレームにはノイズも存在するため、アノテーション作業での運用には工夫が必要である。

アノテーションを行う全てのデータができるだけ正確となるよう運用するのが最も望ましいが、現実的な面言えば、例えば、初期の段階では、全体からサンプルした一部のデータに出現する述語のみ、あるいは、主要語のみに絞って精緻な格フレーム辞書を作りアノテーションを行い、これを評価用のデータとして、残りの部分は、自動獲得した格フレーム等を参考にしながらタグ付けするなどして、大規模にアノテーションするなどの方法で、質と量の双方を兼ね備える方法が好ましいと考えられる。このような方法論は、BCCWJ のコアデータとデータ全体の間関係などにも見られる。

### 5.3.3 非文へのアノテーション

アノテーション対象のデータには、場合によっては一部非文(らしき文)も含まれる。このような文に対して、どのような方針を取るのかについても考慮の余地がある。

(33) a. 服を 乾燥する (受容の余地あり)

b. ガラスを 壊れる (マークされている意味機能的に受容出来ない)

例えば、一般には「乾燥する」は自動詞だとされているが、(33a) のような用例は、Web 上には多数見られる。一方で(33b) の「壊れる」のように、形態論上、自動詞の形を取っているにもかかわらず、格助詞「を」を取るような構造の文は、相対的に受容しがたい。このような文に対して (i) アノテーションするか、ということ、(ii) どのような文を非文とみなすか、ということが問題となる。(i) に関しては、現実データ上に存在する事例であり、応用事例によっては、特によく使われている過ちは頑健に解析したいという場合もあるため、書かれたままの表層格を基にアノテーションする方法が望ましいと思われる。(ii) に関しては、我々の知る限り、現在までに、非文というものの明確な定義は存在せず、個人の内省にもとづいて判断されるもののため、例えば、非文かどうかの判断はアノテータに任せ、代わりに非文と判断された事例を記録しておくことで、必要に応じてデータを区分できるようにしておくような方法が有効な方法として考えられる。

#### 5.4 新聞ドメイン以外で見られた現象

本節では、新聞記事以外のドメインに対する試験的なアノテーション作業において現れた、既存のガイドラインで対象としてない項目についてまとめる。このことについて我々が分析の対象としたデータは、Wikipedia および BCCWJ コアデータ<sup>\*13</sup>より収集した1,000~1,200文程度であるため、各項目の事例を網羅的に収集するに至ったとは言い難い。従って、ここでは、それぞれの項目についての現象の説明をするにとどめ、具体的な考察については、今後の課題とすることにした。

##### 5.4.1 述語の省略

口語的な文においては、文末の述語が省略され、項のみが残されるというケースがよく見られる。

- (34) a. タモリさんから、「これは誰から？」と聞かれた。(「貰ったの」の省略?)

このような場合に、省略されている述語が文脈上容易に想像できる場合、何かしらの述語を補うか、あるいは「述語-非出現」などのラベルを用意して、対応する格をタグ付けするか、そもそも項構造を解析しないか、ということが議論の対象となる。述語省略の究極的なケースとしては

- (35) a. これは...  
b. それはちょっと...

などがあり、このような場合は、述語が何であるかのみならず、残された格が何格であるかすら推定が難しい場合があるため、どこまでがタグ付けして有用な情報かの判断は難しい。

##### 5.4.2 疑問文の照応

対話文においては、疑問文とその回答の間での照応も存在する。

- (36) a. 「あれは誰?」「彼は山田太郎だよ」  
b. 「誰からもらったの?」「太郎からだよ」

NTC・KTCにおいては、現在のところ疑問文に対する照応関係の取り扱いはない。共参照・照応については、本論文での議論の範疇外としたが、対話文の多いドメインに対して照応・述語項構造をタグ付けする場合は、疑問文とその回答に対する、タグ付けのガイドラインも考慮する必要が出てくる。

##### 5.4.3 音象徴語

次の例のように、音象徴語がサ変名詞のように振る舞い、述語として現れる場合がある。そのような場合、音象徴語にも述語項構造をタグ付けしたいが、事例によっては、副詞的振る舞いと、サ変名詞的振る舞いのどちらと取るか判

断に迷う場合があった。

- (37) a. [胸<sub>ga</sub>] がドキドキ する  
b. [胸<sub>ga</sub>] が ドキドキ  
c. [胸<sub>ga</sub>] の ドキドキ (イベント性名詞)  
d. [胸<sub>ga</sub>] がドキドキと 高鳴る 音 (副詞用法)

サ変名詞的振る舞いをする場合、副詞的振る舞いをする場合の他、その他の音象徴語の統語的振る舞いについて、述語として認定するための明確なガイドラインを整備する必要がある。

## 6. 見通しの良いフレームワークの設計

より質の高いアノテーションを目指してガイドラインを改善していくことを考えた場合、対象のガイドラインは、その中で定められたそれぞれの基準がどのような視点で採用されたのかが明確に分かるものでなければならない。また、仕様策定時の理念をコーパス作成者の中で閉じた情報とせず、広く研究者間で共有できる形に整理することにより、継続的な議論が可能になると考える。

このような背景から、本節では、述語項構造アノテーションを題材とすることで集約した、ガイドラインの設計時に配慮されるべき基本方針を述べる。これらは、議論に関わった研究者らが種々のアノテーションタスクの設計を通して経験的に理解している事柄を集約したものであり、複雑でタグ付けコストが高く、また、現象の網羅のために大規模なアノテーションを行う必要がある同様のタスクに対しても有用なものである。以降では、5節の議論から集約した策定上の着眼点を、各論点との対応関係を示しながら述べる。

### データ内の現象に関する取り扱いの網羅性

(5.1.1 節, 5.1.2 節, 5.1.3 節, 5.2.3 節, 5.2.4 節): 大規模なデータに対してアノテーションを施す場合、そのデータ内で起こりうる、判断に特別のガイダンスを必要とする現象に対して、現状のガイドラインがその現象のそれぞれの事例を十分に被覆できるかどうかについて十分な配慮が必要である。特に、アノテーションの判断の決め手が、単語の意味や、慣用表現など、言語の生成的部分に関与している場合には、予め判断基準が列挙しつくせない場合もある。その場合、アノテーション作業中に逐次的にガイドラインの一部を更新・反映する仕組みについても考慮する必要がある。

綿密に判断基準を決めるべき特定の現象があるとき、そのバリエーションが、事前に少量の努力もしくは既知の知識で列挙可能かどうか、非生成的で有限個なのか生成的なのかについて考察し、もし、事前列挙不可能な場合は、作

<sup>\*13</sup> 書籍、雑誌、白書、Yahoo!知恵袋、Yahoo!ブログの5つのドメインを含む。



業中に新たなバリエーションを発見した際に、他の基準に極力影響しない形でガイドラインを更新する方法についても、フレームワークの設計時点で考慮する必要がある。

#### 利用目的とアノテーションスキーマの関係

(5.1.1 節, 5.1.3 節, 5.2.5 節): テキストに対するタグ付けを考えたとき、一般には、利用目的が異なれば、それに応じて必要になるタグ情報も異なる。必要な情報のみを表現するタグセットを作成し、不必要な情報の付与は避けるのが自然なスキーマ設計の方法である。述語項構造のような基本的構造のアノテーションにおいても、同様の考え方は必要である。例えば、次に挙げる応用処理においては、述語項構造のどの部分の情報を用いたいのか、どのような目的で用いるかによって、前提となるタグ付けの基本単位や意味論の精度度が異なる。

- 機械翻訳: フレーズベースで翻訳する場合には、複合語内部の項構造情報は比較的必要性が低い。また、項のラベルについては、基本的には表層格のレベルで十分な場合が多い。日英の場合、省略(ゼロ照応)解析は重要課題とされる。
- 含意関係・言い換え認識: 複合語内部も分解して解析する必要がある。場合によっては、項構造が意味役割のレベルで表現されるのが望ましい。

述語項構造のような基本的な構造をタグ付けする際は、様々な応用の可能性を想定しなければならない。また、このような、それぞれの応用処理からの異なる要求に対し、柔軟にアノテーションの方法を提供でき、かつ、仮にアノテーションがタスク志向で行われた場合にも、最終的にタスク個別のタグ付けデータを統合できるようなフレームワークであることが好ましい。

#### 段階的に質と情報密度を向上できるフレームワーク

(5.2.4 節, 5.2.5 節): 述語項構造や語義のアノテーションのように、アノテーションコストの非常に高いタスクでは、人的資源の制約から、一部の現象のみに対象を絞って初期のアノテーションが行われる場合もある [25]。また、解析のための理論自体が未解決なタスクの場合、全ての問題点が解決するのを待たずに、判断の境界が明確な部分についてのみタグ付けを進めるという方法が取られる場合もある。

こうしたアノテーションタスクにおいては、アノテーションスキーマを順次拡充・変更し、段階的に新しいタイプの情報を付加していける設計のほうが好ましい。しかし、そのような設計の実現のためには、実際にはガイドライン設計の初期段階においても、将来追加される情報をある程度想定し、表現の親和性や、データの自然な統合方法について配慮しておく必要がある。仮に問題を全て解決せずとも、タスク内で起こりうる現象について把握し、可能

な限りスキーマ統合のためのインターフェースを用意しておくのがよい。

#### 本質的に曖昧な選択肢に対する作業の一貫性・評価

(5.2.1 節, 5.3.1 節): ある現象に対してアノテーションを行うとき、複数の選択肢に関して、理論的な要請も特になく、応用処理の観点からもどちらの選択肢を取っても問題ない場合がある。このような場合にどちらの選択肢を選ぶかの基準を設けなければ、その曖昧性のため、見かけ上の作業員間一致率が低下する。また、明確に判断可能な事例と、本質的に曖昧性のある事例が混在することで、解析システムの評価時にも、各事例において、解析誤りのために精度が低いのか、タスク自体が持つ曖昧性のために精度が低いのかの判断が難しくなる。

したがって、このような事例に対して、作業の一貫性を与える、もしくは適切な評価手法を与える配慮が必要である。このような場合、一般には、選択に優先順位のルールを定義することでタグ付けが一意に定まるよう仕様を決める方法が取られる。結果、一致率が向上し、応用処理に用いる際も一貫した利用方法を考えることができる。もう一つの解決策は、本質的に曖昧な事例が出現した際に、その事例が曖昧であることを示しておくことである。そうすることで、曖昧な事例に関しては評価に含めない、あるいはどちらでも正解とみなすなどして、一致率、精度の評価がより適切に行えるようになる。

#### 作業上のコストや作業員が直面する選択肢数をできるだけ減らすスキーマ

(5.1.1 節, 5.1.3 節): 複雑で作業コストが高く、また、表現のバリエーションや頻度分布を観測するために大規模な事例数が必要なタスクにおいては、限られた資源を用いてより多くのデータを作成できるよう、いかにその作業上のコストを下げるかを検討することも重要な課題の一つである。加えて、作業員に複数の選択肢から一つを選ぶような判断を迫る場面においては、できるだけ選択肢を事前に絞り込み、不要な迷いを避ける工夫が、作業効率の面だけでなく、作業結果の一貫性を向上させる意味でも、必要不可欠である [1]。

#### データ量と質のコントロール

(5.3.2): 前述のとおり、大規模なアノテーションを行うためには、一つ一つの事例に対して大きな作業コストのかかる方法を気軽に採用することは難しい。一方で、述語項構造のような基礎的な構造の分析に関しては、解析システムの正確な評価のために、できるだけ高品質なデータが必要とされることも確かである。このような、データ量と品質のトレードオフをどのような方針で管理するかについても、考慮が必要である。



## 7. まとめ

本論文では、より洗練された述語項構造アノテーションのガイドラインを作成する目的で、NTC・KTCの仕様策定、仕様準拠のアノテーション、応用処理に関わった研究者、アノテータらの考察を基に、議論の対象となる点を整理した。具体的には、既存のガイドラインを用いた新規アノテーションによる考察と、研究者・アノテータが経験的に持つ知見を集約するという方法の二つの方法で、既存のガイドラインからは簡潔に解決出来ない問題として4種15項目の論点を洗い出し、それぞれの論点について現状の問題点やそれに対する改善策を議論し、報告した。議論結果を整理するにあたっては、ガイドライン策定の基準となる着眼点を示し、議論内容との対応関係を示すことで、将来のガイドライン改善に向けて建設的な知見となることを目指した。ここに記した問題が、残る問題の全てとは限らないが、こうした建設的な考察の積み重ねによって、実用に耐えうる一貫性を持ったアノテーション方針が作られるとともに、統一かつ頑健な言語解析理論の基礎が積み上がるものと信じるものである。

我々の考察の手順や結果を例に取ると、問題点の洗い出しの方法論や、ガイドライン作成時の理念など、アノテーションに関わる科学は、未だ経験的知見によるところが大きい。しかし、近年では、アノテーションタスクの複雑度や、一致率に影響する因子などに客観的指標を与えようと試みる研究も見られる [1], [4]。6節において、我々が経験的知見によりガイドライン設計の指針としている事柄についても、広く一般的に成り立つ指針として、客観的指標で評価できるような仕組みを生み出していくことも今後の課題である。

## 参考文献

- [1] Bayerl, P. S. and Paul, K. I.: What determines inter-coder agreement in manual annotations? a meta-analytic investigation, *Comput. Linguist.*, Vol. 37, No. 4, pp. 699–725 (2011).
- [2] Burchardt, A. and Pennacchiotti, M.: FATE: a FrameNet-Annotated Corpus for Textual Entailment., *Proceedings of LREC 2008* (2008).
- [3] Carreras, X. and Màrquez, L.: Introduction to the CoNLL-2005 shared task: semantic role labeling, *Proceedings of the Ninth Conference on Computational Natural Language Learning*, pp. 152–164 (2005).
- [4] Fort, K., Nazarenko, A., Rosset, S. et al.: Modeling the complexity of manual annotation tasks: A grid of analysis, *Proceedings of the International Conference on Computational Linguistics (COLING 2012)*, pp. 895–910 (2012).
- [5] グループ・ジャマシイ (編) : 教師と学習者のための日本語文型辞典, くろしお出版 (1998).
- [6] Hashimoto, C. and Kawahara, D.: Construction of an idiom corpus and its application to idiom identification based on WSD incorporating idiom-specific features, *Proceedings of the conference on empirical methods in natural language processing*, Association for Computational Linguistics, pp. 992–1001 (2008).
- [7] Hovy, E., Marcus, M., Palmer, M., Ramshaw, L. and Weischedel, R.: OntoNotes: the 90% solution, *Proceedings of the human language technology conference of the NAACL, Companion Volume: Short Papers*, Association for Computational Linguistics, pp. 57–60 (2006).
- [8] Iida, R., Komachi, M., Inui, K. and Matsumoto, Y.: Annotating a Japanese text corpus with predicate-argument and coreference relations, *Proceedings of the Linguistic Annotation Workshop*, Association for Computational Linguistics, pp. 132–139 (2007).
- [9] Iida, R. and Poesio, M.: A cross-lingual ILP solution to zero anaphora resolution, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pp. 804–813 (2011).
- [10] Kawahara, D. and Kurohashi, S.: Case frame compilation from the web using high-performance computing, *Proceedings of the 5th International Conference on Language Resources and Evaluation*, pp. 1344–1347 (2006).
- [11] Kawahara, D., Kurohashi, S. and Hasida, K.: Construction of a Japanese Relevance-tagged Corpus., *LREC* (2002).
- [12] Kurohashi, S. and Nagao, M.: Kyoto University Text Corpus Project, *Proceedings of the Annual Conference of JSAI*, Vol. 11, pp. 58–61 (online), available from (<http://ci.nii.ac.jp/naid/10011366685/en/>) (1997).
- [13] Laparra, E. and Rigau, G.: Impar: A deterministic algorithm for implicit semantic role labelling, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*, pp. 33–41 (2013).
- [14] Liu, D. and Gildea, D.: Semantic role features for machine translation, *Proceedings of the 23rd International Conference on Computational Linguistics*, Association for Computational Linguistics, pp. 716–724 (2010).
- [15] Loper, E., Yi, S.-T. and Palmer, M.: Combining lexical resources: mapping between propbank and verbnet, *Proceedings of the 7th International Workshop on Computational Linguistics, Tilburg, the Netherlands* (2007).
- [16] Loper, E., Yi, S.-T. and Palmer, M.: Semlink 1.1, , available from (<http://verbs.colorado.edu/semLink/>) (accessed 2013-10-01).
- [17] Màrquez, L., Carreras, X., Litkowski, K. C. and Stevenson, S.: Semantic role labeling: an introduction to the special issue, *Computational linguistics*, Vol. 34, No. 2, pp. 145–159 (2008).
- [18] Matsubayashi, Y., Miyao, Y. and Aizawa, A.: Building Japanese Predicate-argument Structure Corpus using Lexical Conceptual Structure, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)* (2012).
- [19] Matsumoto, Y.: A syntactic account of light verb phenomena in Japanese, *Journal of East Asian Linguistics*, Vol. 5, No. 2, pp. 107–149 (1996).
- [20] 松吉 俊, 佐藤理史, 宇津呂武仁: 日本語機能表現辞書の編纂, 自然言語処理, Vol. 14, No. 5, pp. 123–146 (2007).
- [21] Meyers, A., Reeves, R., Macleod, C., Szekely, R., Zielinska, V., Young, B. and Grishman, R.: The NomBank project: An interim report, *HLT-NAACL 2004 workshop: Frontiers in corpus annotation*, pp. 24–31 (2004).
- [22] Moor, T., Roth, M. and Frank, A.: Predicate-specific Annotations for Implicit Role Binding: Corpus Annotation, Data Analysis and Evaluation Experiments (2013).

- [23] 森田良行, 松木正恵: 日本語表現文型 用例中心・複合辞の意味と用法, *アルク* (1989).
- [24] Palmer, M., Kingsbury, P. and Gildea, D.: The Proposition Bank: An Annotated Corpus of Semantic Roles, *Computational Linguistics*, Vol. 31, No. 1, pp. 71–106 (2005).
- [25] Pradhan, S. S., Loper, E., Dligach, D. and Palmer, M.: SemEval-2007 task 17: English lexical sample, SRL and all words, *Proceedings of the 4th International Workshop on Semantic Evaluations*, Association for Computational Linguistics, pp. 87–92 (2007).
- [26] Ruppenhofer, J., Ellsworth, M., Petruck, M., Johnson, C. and Scheffczyk, J.: FrameNet II: Extended Theory and Practice, *Berkeley FrameNet Release*, Vol. 1 (2006).
- [27] Sasano, R., Kawahara, D., Kurohashi, S. and Okumura, M.: Automatic Knowledge Acquisition for Case Alternation between the Passive and Active Voices in Japanese, *Proceedings of EMNLP 2013 (to appear)* (2013).
- [28] Shen, D. and Lapata, M.: Using Semantic Roles to Improve Question Answering., *EMNLP-CoNLL*, pp. 12–21 (2007).
- [29] Silberer, C. and Frank, A.: Casting implicit role linking as an anaphora resolution task, *Proceedings of the First Joint Conference on Lexical and Computational Semantics - Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, pp. 1–10 (2012).
- [30] Taira, H., Fujita, S. and Nagata, M.: A Japanese predicate argument structure analysis using decision lists, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 523–532 (2008).
- [31] Yoshikawa, K., Asahara, M. and Matsumoto, Y.: Jointly Extracting Japanese Predicate-Argument Relation with Markov Logic, *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pp. 1125–1133 (2011).
- [32] 橋田浩一: GDA 日本語アノテーションマニュアル 草稿 第 0.74 版, , 入手先 (<http://icontent.org/gda/tagman.html>) (参照 2013-10-01)
- [33] 小町 守, 飯田 龍: BCCWJ に対する述語項構造と照応関係のアノテーション, 日本語コーパス平成 22 年度公開ワークショップ, pp. 325–330 (2011).
- [34] 村木新次郎: 日本語動詞の諸相, ひつじ書房 (1991).
- [35] 浅原正幸, 松本裕治: ipadic version 2.6.3 ユーザーズマニュアル, , 入手先 (<http://chasen.naist.jp/stable/doc/ipadic-2.6.3-j.pdf>) (参照 2013-10-01)
- [36] 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, pp. 495–498 (2002).
- [37] 河原大輔, 笹野遼平, 黒橋禎夫, 橋田浩一: 格・省略・共参照タグ付けの基準, , 入手先 (<http://nlp.ist.i.kyoto-u.ac.jp/nl-resource/corpus/KyotoCorpus4.0/doc/rel-guideline.pdf>) (参照 2013-10-01)
- [38] 日本語記述文法研究会 (編): 現代日本語文法, Vol. 2, くろしお出版 (2009).
- [39] 泉 朋子, 今村賢治, 菊井玄一郎, 藤田 篤, 佐藤理史: 正規化を指向した機能動詞表現の述部言い換え, 言語処理学会第 15 回年次大会, pp. 264–267 (2009).
- [40] 林部祐太, 小町 守, 松本裕治, 隅田飛鳥: 日本語テキストに対する述語語義と意味役割のアノテーション, 言語処理学会第 18 回年次大会, pp. 397–400 (2012).
- [41] 佐藤理史: 基本慣用句五種対照表の作成, 情報処理学会研究報告. 自然言語処理研究会報告, Vol. 2007, No. 35, pp. 1–6 (オンライン), 入手先 (<http://ci.nii.ac.jp/naid/110006248060/>) (2007).
- [42] 飯田 龍, 小町 守, 井之上直也, 乾健太郎, 松本裕治: 照応関係タグ付けマニュアル 第 0.02.1 版, , 入手先 ([https://www.cl.cs.titech.ac.jp/~ryu-i/coreference\\_tag.html](https://www.cl.cs.titech.ac.jp/~ryu-i/coreference_tag.html)) (参照 2013-10-01)
- [43] 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).
- [44] 橋本 力, 黒橋禎夫, 河原大輔, 新里圭司, 永田昌明: 構文・照応・評判情報つきブログコーパスの構築, 言語処理学会第 15 回年次大会, pp. 614–617 (2009).