

# 絵本述語項構造コーパスの設計と構築

折田 奈甫<sup>†</sup> 石井 啓太<sup>‡</sup> 鈴木あすみ<sup>§</sup> 松林 優一郎<sup>†</sup>  
 東北大学大学院<sup>†</sup> 情報科学研究科<sup>‡</sup> 国際文化研究科<sup>§</sup> 文学研究科

{naho@ecei, keita.ishii.t1@dc, asumi.suzuki.p7@dc, y-matsu@ecei}.tohoku.ac.jp

## 1 はじめに

本稿では、我々が作成した、日本語の絵本テキストに対して述語項構造タグを付与したコーパスについて紹介する。

言語獲得研究においてインプットの詳細な分析は不可欠である。言語経験から何をどの程度学びうるのか、あるいは学ばないのか、どのような先行知識を仮定すれば学習が可能なのか、というような言語獲得の本質的問題と関係するからである [2]。

日本語のインプットを分析した言語獲得研究では、主に子ども向け発話データベース CHILDES[5] が用いられてきた。日本語の CHILDES では形態素解析が可能であるが、統語・意味・談話についての情報は付与されていない。<sup>1</sup>そのため、これらの言語レベルの問題を検証する際には、個々の研究者が個別の現象に合わせて独自にコーディングする方法が多く取られてきた。また、子ども向け自然発話から多くの知見が得られてきた一方で、自然会話の一部を切り取ったデータであるがために、共有する知識や文脈が推定しにくいことや、視覚情報が伴わないデータが多い<sup>2</sup>などの制限がある。表層の言語情報のみではとらえられない意味・談話レベルの言語獲得の問題をあつかうには、視覚、意味、談話情報などに対して体系的にアノテーションした言語資源が望まれる。

これらのギャップを埋めるために、我々は絵本テキストに対して述語項構造タグを付与したコーパスの構築を行う。絵本には日常会話よりも複雑な文構造が多く含まれており [4]、絵本の読み聞かせは子どもの言語発達を促進すると報告されている [9]。このように子どもの言語経験において重要な役割を果たす絵本は、日常生活の一部を切り取った子ども向け自然発話と比較して、談話や視覚情報（絵本の絵）が限定されている。そのため、視覚情報を含んだ意味関係や談話情報をうまくとらえられる可能性がある。また、述語項構造は、文の構造のみならず、動詞の意味、項省略、格など、言語獲得の主要な問題と密接に関係する。将来的には共参照タグを付与し、述語項構造アノテーションと組み合わせることで、談話レベルの問題にも対応できる。絵本テキストに対して述語項構造をアノテーションすることで、これまでの子ども向け自然発話データでは扱えなかった言語獲得の問題に対応できるコーパスの構築を目指す。

本研究は、日本語の述語項構造アノテーション [13] で培われてきた知識や技術を生かし、以上の特徴と可能性を持つ絵本テキストに対して、以下表 1 のような述語項構造アノテーションを試みる。本アノテーションは、先行する述語項構造アノテーション [13] と比較して以下の特徴を持つ点で新しい。①自動詞、他動詞など動詞の種類のタグを付与する。②述語の主語と補部のみに項タグを付与し、意味役割を担う付加部にはタグを

付与しない。③項に相当する名詞句の指示対象の有生性のタグを付与する。④視覚情報を取り入れた分析を進めるため、絵に指示対象が出ているかどうかのタグを付与する。

表 1: アノテーション見本 (角括弧内は省略を表す)

文	猫が	[鳥に]	お花 [を]	あげた
動詞	NA	NA	NA	3 項
項	項	省略	項	NA
文法関係	主語	間接目的語	目的語	NA
有生性	有生	有生	無生	NA
絵の有無	あり	あり	あり	NA

本稿以下では、コーパスの作成手順、アノテーションガイドラインと、20 冊の絵本 (1174 文) に対してアノテーション作業を行なった結果の作業員間一致率を報告する。

## 2 コーパス作成手順

日本語のミリオンセラーの絵本 (現段階で 85 冊) を選定し、<sup>3</sup> 図 1 の手順でコーパスを作成する。

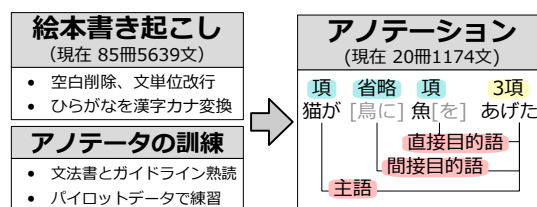


図 1: コーパス作成手順の概要

絵本の本文テキストの人手による書き起こし作業<sup>4</sup>では、アノテーション作業効率化のために、絵本に元々ある、主に文節区切りを表す空白やレイアウトのための改行を削除し、文単位で改行した。この文単位による改行後の行数の合計は 5639 行となった。絵本はひらがなが多く一般的な解析器ではうまく解析できないため [18]、文単位改行後にテキストのひらがなをできるだけ常用漢字やカタカナに変換し、これをアノテーション対象のデータとする。<sup>5</sup>

絵本の読み聞かせ対象年齢は 0 歳から 5 歳の範囲で、3 歳から 4 歳を対象とする絵本が最も多く (47 冊)、次に 0 歳から 1 歳が多い (27 冊)。

<sup>1</sup> 子供向け自然発話のための統語解析器の開発は進められているが、実用には至っていない [12]。

<sup>2</sup> 動画を用いて視覚情報 (指示対象である物体の有無) を含めた日本語の子ども向け発話を分析した研究に、[3] などがある。

<sup>3</sup> 雑誌 MOE2015 年 11 月号を参考にした。

<sup>4</sup> 既存の日本語の絵本データベース [17] は存在するが、外部の研究者が利用できない。

<sup>5</sup> 漢字変換後のテキストに対して形態素解析を行い、動詞の語彙素の自動抽出を試みたが、誤りはごく僅かだった (例: 「渡り」の品詞違い)。

### 3 アノテーションガイドライン

表 1 のように、述語とその項（省略含む）、述語と項の文法関係、項の指示対象の有生性、項の指示対象が絵本の絵に出ているかどうかの 5 種類の情報に対しタグ付けを行う。タグ付与範囲は IPA 品詞体系 [16] の形態素区切りに基づく。

#### 3.1 先行研究との比較

自然言語処理における述語項構造アノテーションでは、意味役割を担う述語の付加部に対しても表層格を用いたタグを付与しているが、意味役割の判断には揺れが多い [15]。この知見から、我々の仕様では、まず第一段階として、述語の主語と補部のみに項のタグを付与し、文法関係のみにアノテーションをする。これらは、本コーパスが対象とするレベルのインプット分析（動詞、格、項省略、照応、談話など）に共通して最低限必要であると判断した情報である。

また、これまでの日本語の述語項構造アノテーションでは、述語と項の関係に対して表層格を用いてタグ付けをしているが、我々の仕様では表層格を用いず、文法関係タグを付与する。これは以下の例 (1) のように、表層格と文法関係が必ずしも一致しないためである [14]。

(1) 花子に <sub>二格主語</sub> この問題が <sub>ガ格目的語</sub> 解ける。

これまでの述語項構造アノテーションや子ども向け自然発話データベースにはない新しいタグとして、項に相当する名詞句の指示対象の有生性と、これらの指示対象が絵に出ているかどうかのタグを付与する。有生性は、述語の構造・意味 [11] や指示対象の顕著性 [7] と関係するため、表層の言語情報からは特定できないインプットの分布を分析するために有用だと考えた。また、視覚情報は、子どもの指示表現選択 [7] に影響すると報告されており、項省略や談話レベルの分析に新たな知見を供する可能性があると考えた。言語獲得の問題を検証したコーパス研究は、言語情報のみを分析対象とすることが多いが、実際に子どもが言葉を学習する状況では視覚などの多様で豊かな情報が利用可能である。今回のアノテーションでは指示対象が絵に出ているかという単純なタグに限定するが、将来的にはより粒度の高い視覚情報を扱いたい。

#### 3.2 各アノテーション項目の説明

**述語** 自動詞、他動詞、3 項動詞、形容詞、コンピュータ、コンピュータ省略\*6のどれかのタグを付与する。他動性は連続的であるため [11]、自動詞と他動詞の判断が困難な場合があれば、[11] を参考に、その動詞が意味的に他の事物に影響を及ぼすかどうか、受動文や再帰文などに変換できるかどうかなどにに基づき判断する。省略された述語についてはタグを付与していない。

**項** 述語の項に付与されるタグには、項、補文、イベント、追加項の 4 種類がある。全ての述語の主語に項のタグを付与し、他動詞は目的語、3 項動詞は直接目的語と間接目的語に項のタグを付与する。形容詞とコンピュータは基本主語のみだが、「ほしい」「好き」「必要」などの形容詞・形容動詞については、目的語に相当する句に対しても項のタグを付与する。項のタグが付与されるのは、主語に相当する句と補部に相当する句（目的語/間接目的語）のみで、付加部に相当する句には項のタグを付与しない。項に相当する名詞句のタグ付け範囲は最大名詞句としてい

\*6 コンピュータとコンピュータ省略には形容動詞文も含む。

る。また、補文に相当するものには、「と」「の」「か」などの補文標識に対して補文タグを付与し、「騒ぎ」「試合」など、実体を指示せずイベントや抽象的な概念を指示する名詞句にはイベントのタグを付与する。

1 人称、2 人称、省略項については、先行する述語項構造アノテーション [13] にならい、以下 (2) ように専用の文字列 (**exo1**: 1 人称、**exo2**: 2 人称、**null**: 省略) を各文頭に用意し、これらに対して項タグを付与する。

(2) a. **exo1**<sub>項</sub> **exo2** **null** 「どんぐり拾ったよ」(発話者)

b. **exo1** **exo2**<sub>項</sub> **null** 早く寝なさい! (命令文の主語)

c. **exo1** **exo2** **null**<sub>項</sub> 魚を食べたみたいだ。(主語省略)

同じ文中の異なる動詞句や節に、指示対象を示す名詞句が明示的に出ている場合には、その名詞句にタグを付与する (例 (3a): 「兎」は「つく」と「焼く」の両方の主語としてタグ付けされる)。ただし、省略項と明示的に出ている名詞句の文法関係が異なる場合は、その名詞句にタグを付与せず、別に省略 **null** にタグを付与する (例 (3b): 「熊が持って来た食べ物」(どんぐり) は「見る」の目的語であり、後続するコンピュータの主語である。文法関係が違うためコンピュータの主語は **null** になる)。

(3) a. 兎 <sub>項:主語</sub> はお餅をついて、 $\phi$  焼きました。

b. **null**<sub>項:主語</sub> 熊が持ってきた食べ物 <sub>項:目的語</sub> を見ると、 $\phi$  どんぐりだった。

**文法関係** 述語とそれぞれの項の文法関係（主語、目的語、間接目的語）を付与する。先行する述語項構造アノテーション [13] では、意味役割を担う述語の付加部に対しても表層格を用いてタグを付与しているが、3.1 で述べたように、このアノテーションでは意味役割は考慮せず、文法関係に限定する。意味役割を扱わないため、動詞の種類によってタグ付けされる項の数と文法関係が決まる仕様となっている。副助詞「は」「も」などは、ガ格ヲ格と表層で共起しないため、副助詞の付く名詞句に関しては、先行研究のテスト [6] を参考に、文法関係を担う項なのか、それとも主題などの情報構造を示すだけなのかを区別している。また、受動、使役、可能などの態や授受表現が使われると、項が追加されて表層格と文法関係が異なる場合がある。この場合は、(4) のように述語に態のラベルを付与し、追加項以外の項に文法関係を付与する。(4) の追加項は使役文の主語に相当するが、この文法関係は付与していない。

(4) 母 <sub>追加項</sub> が娘 <sub>項:主語</sub> に本 <sub>項:目的語</sub> を読ま <sub>他動詞:使役</sub> せた。

**有生性と絵の有無** 項には有生性のタグ（有生、無生、不明、該当なし）と、項の指示対象が絵に出ているかどうかのタグ（絵にあり、絵になし）を付与する。有生性は、その個体に心や意識があり動くかなどを考慮して判断する。絵の有無は、文と同一ページにある絵のみを対象に判断する。

## 4 作業員間一致率

### 4.1 作業手順

今回一致率を報告するデータのアンノテータは、言語学の基礎的訓練を受けた言語学研究室の大学院生二名（共著者）である。アンノテータは、基本的な日本語の文法知識の復習 [10, 19] とガイドラインの熟読を行い、主著者からアノテーション方法について指導を受けた後、パイロット用データにアノテーションし、主著者がその結果を確認するという手続きを取った。しかし、

今回一致率を測ったデータにおいても議論すべき点が多くあったため、本稿で報告する結果は、3回のガイドライン改定を反映したものである。タグ付けは、アノテーションツール brat[8]を用いて行った。

## 4.2 作業員間一致

上述した二名のアノテータが20冊1174文の絵本にアノテートした結果の作業員間一致を表2に示す。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、適合率、再現率、F値を計算した。

**述語** 述語全体の一致率については、動詞の種類とアノテーション範囲が全て一致する場合、種類と述語の開始位置のみが一致する場合、述語であれば種類は問わず開始位置のみ一致する場合の3通りで算出した。述語の種類的一致を考慮する場合(90.8%)と比較して、述語であれば種類は問わない場合のF値は97%以上に達する。この差は、動詞をカテゴリカルに分類することの難しさを反映していると考えられる。自動詞と他動詞の間で揺れがあったもののうち、典型的な例を以下に挙げる。

(5) a. 猫はお家に入った。

b. 馬は草原を超えた。

例(5)のように、移動を表す動詞の目的地や経路を目的語とするか付加部とするかにより、述語ラベルに揺れが出る。例(5a)のような「に」格が付く名詞句が目的語なのか後置詞なのかは、研究者により見解が分かれるところである[19]。例(5b)のような「を」格が付く名詞句は、直接目的語ではなく「超える」は自動詞だという分析が言語理論では一般的だが[19]、このような言語学的知識を持つアノテータであっても、実際のアノテーションでは項を要求する他動詞であると判断する傾向がある。

また、以下(6)のように、形態的な対応のある他動詞と3項動詞において、他動詞から使役動詞になったものと3項動詞の間で揺れが見られた(他に「乗せる/乗らせる」など)。このような対応のある動詞は、意味的にも構造的にも類似するため、使役の言語学的知識を持っていても、実際のアノテーションでは混乱する場面が観察された。

(6) a. 猫が鳥に服を着せ<sub>3項</sub>た/着させ<sub>他動詞:使役</sub>た。

b. 猫が鳥に絵を見せ<sub>3項</sub>た/見させ<sub>他動詞:使役</sub>た。

述語の中では3項動詞の一致率が最も低い。これは以下の例(7)のように、「に」格の付く名詞句を、間接目的語と判断するか後置詞と判断するかで揺れが生じる場合が多いためである。例(7a)では、「に」格の付く名詞句を間接目的語と判断し、「かける」を3項動詞とタグ付けする場合と、「に」格の付く名詞句を後置詞として判断し、「かける」を他動詞とタグ付けする場合で揺れがあった(他に「入れる」「突っ込む」など)。例(7b)では、被依頼主である省略項を間接目的語とするか付加部とするかで揺れがあった。

(7) a. 猫がマントを壁にかけた。

b. 「ケーキをください」と熊は $\phi$ 頼んだ。

コンピュータ省略の一致率も低い。本アノテーションでは、省略された動詞はタグ付けしない方針だが、コンピュータの省略だけはタグ付けをしている。このため、例(8)のように、動詞が省略されているか、コンピュータが省略されているかの判断が分かれる場面が観察された。例(8a)では、「いい匂い(だ)」のようにコンピュータが省略されていると判断する場合と「いい匂い(がする)」のように動詞が省略されていると判断する場合で揺れが

あった。同様に例(8b)では、「次はマラソン(だ)」のコンピュータ省略と「次はマラソン(しよう)」の動詞の省略で揺れがあった。

(8) a. いい匂い $\phi$

b. 次はマラソン $\phi$

**項** 項の一致では、述語との関係は考慮せず、項としてタグ付けされたもの全ての範囲一致をみている。補文、イベントを表す名詞句、態により生じる追加項の一致率もそれぞれ算出した。

イベントを表す名詞句と追加項の一致率が低い。イベントを表す名詞句の一致率については、「お詫び」「騒ぎ」のような動詞由来のものだけでなく、「パーティー」のような名詞句もイベントを表す名詞句とみなすかどうかで揺れが観察された。また、以下の例(9)のように、省略されている場合の判断(省略項を「狼」と解釈するか「攻撃」と解釈するか)や、「あっかんべー」を「あっかんべーしている顔」と解釈するか「あっかんべー」という動作であると解釈するかの間で揺れがあった。

(9) a. 熊は狼の前に立ちふさがり、 $\phi$ 必死に防いでいた。

b. わー! あっかんべーだ!

追加項の不一致は、他動詞と3項動詞の間の揺れに起因するものが多い。イベントと追加項はともに事例数が少ないため、原因の特定にはさらに事例を収集し分析する必要がある。

**文法関係** 文法関係は、述語の種類と範囲が一致したものを対象に一致率を算出した。主語と目的語ともに省略(1人称と2人称を含む)の一致率が若干低い。ガイドラインでは、省略であっても同じ文中の異なる動詞句や節に明示的に指示対象を指示する名詞句が出ている場合には、その名詞句にタグを付与し、省略項と明示的に出ている名詞句の文法関係が異なる場合には別に省略タグを付与するとしているが、この判断で混乱する場面が観察された。例えば(10a)の省略されている受動態の主語を「木」とするか省略タグを付与するか、(10b)の省略されているコンピュータの主語 $\phi_k$ を「連れて来た猫」とするか省略とするかのような判断で揺れがあった。

(10) a. 木<sub>i</sub>は歳を取り、 $\phi_i$ 新しいのに植え替えられた。

b.  $\phi_i$   $\phi_j$  連れて来た猫<sub>k</sub>を見ると、 $\phi_k$  痩せた野良猫だ。

文法関係がなんでもあれ同じ文中に明示的な名詞句があればそれを項とするというルールに単純化の方がよいと思われる。

**有生性と絵の有無** 項の範囲が一致しているものを対象に、有生性と絵の有無の一致率を算出した結果、有生性は93.3%、絵の有無は92.1%となった。

絵本では、機関車や果物など、本来無生物のものが心を持ち動くため、有生性の判断が分かれる場合がある。また、以下(11)のように、省略された項の解釈の違いが有生性と絵の有無の不一致に反映されている。例(11a)は、猫が消防士たちが火を消すの手伝ったという意味の文であるが、「手伝う」の省略された目的語を「消防士たち」と解釈するのか「火を消す」というイベントと解釈するのかが、有生性と絵の有無の判断が異なる。例(11b)では、「面白い」の省略された主語を、「変な顔をする」というイベントと解釈するか、「変な顔をする人」と解釈するかで、有生性と絵の有無の判断が異なる。

(11) a. 猫は $\phi$ 手伝った。

b.  $\phi$  わー面白い!

このように、有生性と絵の有無の不一致は、省略項の解釈の違いが原因である場合が多い。どちらの解釈も可能な場合が多いため、この一致率は妥当だと考える。

表 2: 絵本テキスト 20 冊分に対する述語項構造アノテーションの作業者間一致

種類	適合率 (%)	再現率 (%)	F1 (%)
述語 - 全て (種類と範囲が一致)	88.4 (1330 / 1504)	87.8 (1330 / 1515)	88.1
述語 - 全て (種類と開始位置が一致)	91.2 (1371 / 1504)	90.5 (1371 / 1515)	90.8
述語 - 全て (述語であれば種類問わず; 開始位置が一致)	99.3 (1452 / 1462)	95.8 (1452 / 1515)	97.6
項 (述語との関係は見ず範囲のみ一致)	94.3 (1519 / 1611)	93.4 (1519 / 1626)	93.9
補文	92.8 (116 / 125)	96.7 (116 / 120)	94.7
イベントを表す名詞句	77.5 (31 / 40)	57.4 (31 / 54)	66.0
追加項	81.3 (13 / 16)	61.9 (13 / 21)	70.3
自動詞 (以下全て範囲一致)	87.3 (543 / 622)	88.0 (543 / 617)	87.7
他動詞	91.4 (469 / 513)	88.2 (469 / 532)	89.8
3項動詞	76.7 (23 / 30)	69.7 (23 / 33)	73.0
形容詞	90.7 (107 / 118)	91.5 (107 / 117)	91.1
コピュラ	86.1 (99 / 115)	89.2 (99 / 111)	87.6
コピュラ省略	84.0 (89 / 106)	84.8 (89 / 105)	84.4
主語 - 全て (以下全て種類と範囲が一致した述語の項)	96.7 (1285 / 1329)	96.6 (1285 / 1330)	96.7
主語 - 省略なし	97.0 (1037 / 1069)	97.7 (1037 / 1061)	97.4
主語 - 省略	95.4 (248 / 260)	92.2 (248 / 269)	93.8
目的語 - 全て	98.1 (504 / 514)	97.9 (504 / 515)	98.0
目的語 - 省略なし	98.3 (403 / 410)	98.5 (403 / 409)	98.4
目的語 - 省略	97.1 (101 / 104)	95.3 (101 / 106)	96.2
間接目的語 - 全て	100.0 (23 / 23)	100.0 (23 / 23)	100.0
間接目的語 - 省略なし	100.0 (19 / 19)	100.0 (19 / 19)	100.0
間接目的語 - 省略	100.0 (4 / 4)	100.0 (4 / 4)	100.0

## 5 結論

本稿では、日本語の絵本テキストに対して述語項構造タグを付与したコーパスを構築するためのアノテーション仕様と作業者間一致率を報告した。

自然言語処理における述語項構造アノテーションの知見や技術を参考にすることで、言語獲得分野では前例のない言語資源の構築であるにも関わらず、見通しよく体系的な仕様の設計が可能となった。作業者間一致では、他動性の判断、「に」格の判断、省略項の判断などに作業者間の揺れを観察したが、これは自然言語の持つ本質的な難しさから生じるものである。言語理論と自然言語処理の知見をさらに取り入れることで、より明確なガイドラインの策定を目指したい。

本プロジェクトの今後の予定として、最終的に絵本 100 冊程度の述語項構造アノテーションを目指している。また、意味役割や、粒度の高い視覚情報のアノテーションも検討したい。述語項構造アノテーションの後には、共参照・照応のアノテーションを予定している。これにより、顕著性、話題性、情報構造など、様々な言語現象と関連する談話レベルの情報 [1] を定量化できるようにし、インプット分析や学習シミュレーションに役立てたい。

## 謝辞

統語論の観点から適切な助言をくださった東北大学の高橋大厚先生と、有益なフィードバックをくださったアノテータの方々に感謝いたします。本研究は文部科学省科研費 (16H06623) から支援を受けて行われました。

## 参考文献

- [1] Jennifer E Arnold, Elsi Kaiser, Jason M Kahn, and Lucy K Kim. Information structure: linguistic, cognitive, and processing approaches. *WIREs Cognitive Science*, Vol. 4, No. 4, pp. 403–413, 2013.
- [2] Noam Chomsky. *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group, 1986.
- [3] Anne Fernald and Hiromi Morikawa. Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child development*, Vol. 64, No. 3, pp. 637–656, 1993.
- [4] Erika Hoff-Ginsberg. Mother-child conversation in different social classes and communicative settings. *Child development*, Vol. 62, No. 4, pp. 782–796, 1991.
- [5] Brian MacWhinney. The CHILDES project: Tools for analyzing talk, 2000.
- [6] Hiromi Morikawa. *Acquisition of case marking and argument structures in Japanese*. くろしお出版, 1997.
- [7] Ludovica Serratrice. The role of discourse and perceptual cues in the choice of referential expressions in English preschoolers, school-age children, and adults. *Language Learning and Development*, Vol. 4, No. 4, pp. 309–332, 2008.
- [8] Pontus Stenetorp, Sampo Pyysalo, Goran Topic, Tomoko Ohta, Sophia Ananiadou, and Junichi Tsujii. brat: a web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations Session at EAACL 2012*, 2012.
- [9] Grover J Whitehurst, Francine L Falco, Christopher J Lonigan, Janet E Fischel, Barbara D DeBaryshe, Marta C Valdez-Menchaca, and Marie Caulfield. Accelerating language development through picture book reading. *Developmental psychology*, Vol. 24, No. 4, p. 552, 1988.
- [10] 益岡隆志, 田窪行則. 基礎日本語文法. くろしお出版, 1992.
- [11] 角田太作. 世界の言語と日本語 改訂版 言語類型論から見た日本語. くろしお出版, 2009.
- [12] 宮田 Susanne, 寒河江健二, マクウィニー・ブライアン. CHILDES 用の構文解析プログラム GRASP. 健康医療科学研究, No. 3, pp. 45–62, 2013.
- [13] 笹野達平, 飯田龍, 奥村学 (監修). 文脈解析-述語項構造・照応・談話構造の解析-. コロナ社, 2017.
- [14] 柴谷方良. 日本語の分析. 大修館書店, 1978.
- [15] 松林優一郎, 飯田龍, 笹野達平, 横野光, 松吉俊, 藤田篤, 宮尾祐介, 乾健太郎. 日本語文章に対する述語項構造アノテーション仕様の考察. 自然言語処理, Vol. 21, No. 2, pp. 333–377, 2014.
- [16] 情報処理振興事業協会技術センター. 計算機用日本語基本動詞辞書 ipal, 1996.
- [17] 藤田早苗, 服部正嗣, 小林哲生, 奥村優子, 青山一生. 絵本検索システム「びたりえ」～子どもにぴったりの絵本を見つけます～. 自然言語処理, Vol. 24, No. 1, pp. 49–73, 2017.
- [18] 藤田早苗, 平博順, 小林哲生, 田中貴秋. 絵本のテキストを対象とした形態素解析. 自然言語処理, Vol. 21, No. 3, pp. 515–539, 2014.
- [19] 鈴木孝明. 日本語文法ファイル 日本語学と言語学からのアプローチ. くろしお出版, 2015.