様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　４ 年　 ６ 月 １６ 日現在

機関番号：１１３０１
研究種目：基盤研究(C)（一般）
研究期間：2019～2021
課題番号：１９Ｋ００５９３

研究課題名（和文）日本語ツリーバンク文法情報の精緻化

研究課題名（英文）Development of a Japanese treebank with more precise grammatical information

研究代表者
吉本　啓（Yoshimoto, Kei）

東北大学・高度教養教育・学生支援機構・名誉教授

研究者番号：５０２８２０１７

交付決定額（研究期間全体）：（直接経費）　　2,600,000円

研究成果の概要（和文）：研究代表者らは，日本語として初めての統語解析情報付きコーパス NINJAL Parsed Corpus of Modern Japanese (NPCMJ) の開発に携わってきた。同コーパスをさらに充実したものとするために，以下の研究を行った。第一に，従属節の主語が明示されない場合の解釈について，節の種類によってそれぞれ実情に合ったディフォールト規則を設定し，主節の項を継承するものとした。第二に，これまで欠落していたテンス情報のアノテーションをダイナミック意味論にもとづいて施した。第三に，正確な意味解析のために，ディフォールトにもとづくスコープ階層を提案した。

研究成果の学術的意義や社会的意義
従来，日本語については形態素情報を中心とするコーパスしか存在せず，日本語文法研究への本格的な利用は難しかった。研究代表者らの構築した統語解析情報付きコーパス NPCMJ により，構文研究へのコーパス利用の道を開くことができた。今回の研究のうち，従属節の非明示的主語およびスコープのディフォールト解釈は，正確な意味分析を行うためのアノテーション作業を効率化することを可能にする。また，言語的なテンス意味をアノテーションとして施すのは日本語として初めての試みであり，言語学のみならず言語処理においても貴重なデータを提供する。

研究成果の概要（英文）：We have aimed at improving the annotation of NPCMJ, the first Japanese corpus with syntactic information. First, we have set up default rules to interpret an omitted subject within an embedded clause by inheritance from the matrix clause in different ways dependent on the sort of the clause. Second, we have given annotation on tense information which lacked so far. Third, we have proposed a scope hierarchy which works as a default rule for scope interpretation.

研究分野：コーパス言語学

キーワード：コーパス　ツリーバンク　統語論　意味論　日本語

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

現代の言語学において，コーパスは研究の重要な手段として市民権を得ている。日本語についても「現代日本語書き言葉均衡コーパス (BCCWJ)」や「日本語話し言葉コーパス (CSJ)」がある。しかし，これらはテクスト中の文を文節へと分析し，それぞれの単語に形態素情報をタグ付けしたものである。そのため，語彙や形態論の研究においては威力を発揮するが，日本語研究者の関心はそれだけに限られない。文法に関心を持つ研究者にとってこれらのコーパスは限定的な意義しか持たない。

このことから、研究代表者たちは平成 28 年度より、国立国語研究所共同研究プロジェクトとして統語・意味解析情報をタグ付けしたコーパス NINJAL Parsed Corpus of Modern Japanese (NPCMJ) の開発を開始した。現在，約 10 万文のアノテーションをウェブサイトで公開しており、日本語文法研究の基礎的ツールとして認知されつつある。

しかしながら，コーパスが実際に使われるようになって、問題点も明らかになってきている。特に、言語の理論研究の観点からは、より精緻な文法情報を必要とする。その主要なものには、以下の諸点が含まれる。

i. 日本語の従属節において主語が明示されないことがしばしばあるが，これは主節の主題や主語に関する情報が継承されているためと考えられる。このことを，実際の言語データをもっともよく説明できる形で，いかにアノテーションに反映させるか。
ii. 現状の NPCMJ におけるテンス情報としては，助動詞 (AXD) 「た」が存在するか否かが記されているにすぎない。実際のテンス解釈を (単文も複文も含めて) 反映するアノテーションを実現可能な形でいかに行うか。
iii. 正確な文の意味解析情報の提供のためには、正確なスコープ (作用域) の把握が欠かせない。技術的に可能な範囲内で、スコープ表示に関してどのような改善をするべきか。

２．研究の目的

これまでに達成した NPCMJ のアノテーションを基礎として、より精緻で研究上価値の高いアノテーションを行うための方法を確立することを目的とする。具体的には、(i) 従属節の主語が明示されず主節の主題/主語から継承される条件、(ii) テンス情報のアノテーション、および (iii) スコープの柔軟な表示、の 3 つの課題に取り組む。(i)〜(iii) の検討の結果は、アノテーションの改善という形で具体化することが出来る。ディフォールトの意味解釈 (矛盾が生じない限り成り立つと仮定される解釈) を決定して与え、それと一致しない場合のみ明示的にアノテートすればよい。実際に NPCMJ の一部にアノテーションを施し、理論の検証を行う。

３．研究の方法

理論的検討は吉本が行った。森が大学院生の協力によって，完成しつつあった NPCMJ の言語データを検索，分析し，吉本に提供した。

また，科研費新学術領域研究「時間生成学 時を生み出すこころの仕組み」計画研究「言語による時間生成」への参加により，多くのヒントを得ることができた。

４．研究成果

I．従属節の主語の主節からの継承

南（1974）は，文階層構造説の一環として，ある種の従属節では主語や主題が明示されなくとも，主節の主語や主題をディフォールトとして継承すると考えている。この考えに従って NPCMJ のデータ中の従属節の主語と主節の主語や目的語との関係を調査し，従属節中で主語が表現されていない場合の解釈としてもっとも妥当な，以下のディフォールト規則を抽出して，アノテーションに適用した。。

準主節（IP-SUB とタグ付けされる），関係節（IP-REL），および等位接続節（IP-ADV-CONJ）の中の非明示の主語と上位の節の項との間にコントロール関係が成り立つことはない。これに対し，従属節（IP-ADV-SCON(-CND)），空所なし名詞修飾節（IP-EMB），名詞化節（IP-NMZ），および小節（IP-SMC）についてはコントロール関係が成り立つ。

上位節のコントロール元（先行詞）として複数の可能性がある時，指示対象が何であるかは先行詞の文法役割にもとづいて決定される。可能性の高い順に，以下のようになる。

OB2 > OB1 > SBJ2 > SBJ

このようなアクセス可能性の階層に対する例外として，主語のみを先行詞とするコントロール関係がある。このような場合，節のラベルを IP-ADV2-SCON(-CND) とすると，先行詞を上位節の主語に限定することができる。他にも，IP-EMB2，IP-NMZ2,IP-SMC2 とすることにより，コントロールを上位節の主語によるものに限定することができる。

ほとんどのコントロール環境においては，主語の役割以外の先行詞は，コントロールの受け手に対して先行することが条件である。ただし，小節（IP-SMC）については，この条件は課されず，先行詞は後続してもよい。また，先行詞が主語である場合，コントロールの受け手に後続す

ることが一般に許される。

II．テンス情報のアノテーション

上記のように，研究の端緒は複文におけるテンス情報の継承を研究することにあった。しかし，そのためには複文に関わるテンス解釈の問題を日本語のテンス体系全体の中に位置づける必要が生じる。そのため，まず単文をも含む日本語文のテンス情報のアノテーションについての考察が主たる研究成果となった。

提案したアノテーションの特徴として，第一に言語の論理的意味の解析にもとづくものであること，第二にイベント意味論にもとづくものであること，第三にダイナミック意味論を採用していることが挙げられる。

第一に，本アノテーションは，言語的意味というよりもむしろ物理的な時間の論理的関係を分類して推論体系を構築した Allen (1983) やそれにもとづいてコーパスを構築した Pustejovsky (2017) や吉川・浅原 (2015) とは一線を画する。この体系に従うと，2 つの事象の間に少しでも隙間が空いているか，それともぴったりと隙間無く隣接しているかが厳密に区別されるが，そのような区別は言語的な根拠に乏しい。一種の不正確さを許容することによって成立する言語的意味を対象としてこそ言語使用者にとって意味があるし，また均質的なアノテーションが可能になる。

第二に，本アノテーションはイベント意味論にもとづいており，Kamp and Reyle (1993) に従って，イベントを出来事や状態の生起する時間を位置づけるための変項として利用する。

第三に，ダイナミック意味論の一種であるディスコース表示理論 (DRT; Kamp and Reyle 1993) のテンス・アスペクト解析に概ね従っている。ダイナミック意味論/DRT は文脈的意味を取り扱えるという利点を持ち，複数の文にまたがるテンス情報の記述を可能にする。

アノテーションは，以下の原則に従って行う。

文の表す時間的意味を表示するために，出来事および状態を表す変項 (DRT の用語では談話標識) e および s によって事象の成立する時間を表すと同時に，様々な時間関連表現が表す条件の規定にも利用する。この他，時間を表す談話標識として，主として副詞句が表示する時間位置を表す t，発話時を表す n，および参照時を表す r がある。参照時によって，文脈的・語用論的要因を導入することができる。また，時間関係を表すオペレータを表 1 に示す。

Kamp and Reyle (1993) に従い，これら 5 種類の談話標識と 5 つのオペレータによって，日本語テクストにあらわれるすべての時間情報を記述する。Allen (1983) に従うよりも格段に簡潔であり，無用の曖昧性を避けることができる。

テンスに関する規定は，動詞，形容詞，助動詞 (以上，「用言」)，副詞，および接続詞 (従属節ヘッド) の持つ語彙的情報によって導入される。

テンスに関するもっとも基本的な情報は，述語が持つ以下の規定によりもたらされる。

表 1: 時間関係を表すオペレータ

先行/後続	$<$
包含	$\subseteq$
重複	$\bigcirc$
同一	$=$
隣接	$\supset\subset$

(1) 動作述語タ形: $t < n, e \subseteq t$
動作述語非タ形: $n < t, e \subseteq t$
状態述語タ形: $t < n, s \bigcirc t$
状態述語非タ形: $t = n, s \bigcirc t$

以上は，タ形/非タ形の用言主要部とテンス助動詞 (非タ形の場合は，ゼロ語形の助動詞とするか，あるいは用言主要部にすべてを規定) との語彙的情報の組み合わせとして表現できるが，複雑になるので詳細は省く。

これにより，単純な文がどのようなテンス上の解析を与えられるかを以下に示す。(2a) の DRT による解析結果が (2b) である。その中で，上方の行は出現する談話標識のリストであり，述語論理式の量化子にほぼ相当する。下方の行は談話標識の間に成立する条件である。直接テンスに関係する情報のみを記す。(3b) 以降は条件のみを記すことにする。

(2) a. 先月太郎は旅行した。
b.

t e n
先月(t), 旅行する(e), t < n, e ⊆ t

先月(t)，旅行する(e)，t < n，e ⊆ t

(3) a. 来月太郎は旅行する。
b. 来月(t)，旅行する(e)，n < t，e ⊆ t

(4) a. 昨日は寒かった。
b. 昨日(t)，寒い(s)，t < n，s ○ t

(5) a. 今日は寒い。
b. 今日(t)，寒い(s)， t = n，s ○ t

上に述べたように，e および s は事態の発生時を表すが，発話時 n とは直接関係づけられず，時間位置 t を介して関係づけられる。例えば (2b) では， t が n に先行し (t < n)，しかも e は t に包含される (e ⊆ t) ので，結果的に e は n 以前，すなわち過去に発生することになる。また，時間副詞のあらわれない文についても，解析の一貫性を保つために，時間位置を仮定する。

以上では 1 つの文の中に事態や時間位置を表す談話標識が各々1 個しかあらわれない例のみを扱った。しかし，アノテーションで対象とする言語データの中には 1 つの文中に複数の事態や時間副詞が出現するものも多い。それらの中から複文を取り上げて，アノテーションをいかに行ったかを以下に示す。

複文の場合は，南 (1974) による従属節の階層 (A，B，C 類) に応じてテンス解釈に違いが見られることが知られている。

A 類従属節は独自のテンス解釈を持たず，主節のテンスにもとづいて解釈される。この類に属する従属節ヘッド接続詞「ながら」が導く従属節を持つ文のテンス規定は以下のとおりとなる。複文のテンス解釈は，統語構造や主節・従属節中の用言やテンス助動詞 (それが存在する場合には) が持つテンス情報が複雑に統合されて行われるが，本稿ではその結果のみを示す。

(6) ながら: t1 = t2

ナガラ従属節の中に出現する動詞は，テイル形と同一の意味を持つと見なす。主節および従属節の時間位置 t1 と t2 は同一だとされている。これにより，(7a) の解釈は (7b) のように与えられる。

(7) a. 太郎は新聞を読みながらご飯を食べた。
b. 食べる(e1)，Prog(s2，読む)，t1 < n，e1 ⊆ t1，t1 = t2，s2 ○ t2

t1 と t2 は (6) により同一となるので，結果として従属節述語「読み」の発生時 s2 は主節述語「食べた」のそれと重複する (s2 ○ e1)。これにより，主節・従属節述語の発生時の間の関係が説明される。

B 類従属節のテンス解釈は，主節・従属節のテンス表示 (非タ形/タ形) および従属節の述語意義特徴 (動作/状態) に依存して行われる。これもまた，統語構造および各語彙の持つテンス情報の複雑な統合により行われるが，結果のみを表 2 に示す (主節および従属節の時間位置をそれぞれ t_m および t_s とする)。

以上にもとづき，例文 (8a) および (9a) のテンス解釈は (8b)，(9b) のように与えられる。

表 2: B 類従属節を持つ複文のテンス解釈

		従属節: 動作	
		非タ	タ
主節	非タ	$n < t_s$	$t_s < n$
	タ	$t_m < t_s$	$t_s < t_m$

		従属節: 状態	
		非タ	タ
主節	非タ	$t_s = n$	$t_s < n$
	タ	$t_s = t_m$	$t_s < t_m$

(8) a.春子が卒業するので一緒に旅行する。
　　b.旅行する(e1)，卒業する(e2)，n ＜ t1，e1 ⊆ t1，n ＜ t2，e2 ⊆ t2
(9) a.春子が病気だったので電話した。
　　b.電話する(e1)，病気だ(s2)，t1 ＜ n，e1 ⊆ t1，t2 ＜ t1，s2 ○ t2

主節が非タ形の場合，(8b) におけるように，従属節述語の発生時は発話時 n とのみ関係づけられ，主節述語の発生時との関係は未指定である。このことは，(8a) において「卒業する」と「旅行する」の発生時のうちどちらが先行してもかまわないことに対応する。これに対して，(9b) のように主節がタ形の場合，従属節述語と発話時との関係が指定される。
C類従属節を持つ複文では，主節・従属節間でテンス解釈は互いに無関係に，独立して行われる。例文とその解釈を (10a，b) に示す。

(10)a.春子はスペインへ行ったが，明子はイタリアへ行く。
　　b.行く(e1)，行く(e2)，n ＜ t1，t2 ＜ n，e1 ⊆ t1，e2 ⊆ t2

III．スコープの扱い

NPCMJ の言語データの検索・分析を通じて，以下の結論に達した。
語順による優先順位、すなわち先行する修飾部が他よりも広いスコープを取るというディフォールトを基本的原則とする。
述語は，それよりもさらにスコープの階層の低い（スコープの狭い）補部に次いで，もっとも低くなる。また，副詞句 (ADVP) のスコープは，イベントが束縛される範囲よりも狭くなければならない。
主題 (TPC) および呼び掛け (VOC) は，もっとも広いスコープを取る。また，主語 (SBJ) は他の修飾部よりも広いスコープを取る。
スコープの階層は，全体として以下のように与えられる。

HIGHEST とタグ付けされた修飾部
HIGH とタグ付けされた修飾部
SBJ とタグ付けされ、スコープに関してタグ付けされない項
その他の修飾部のディフォールト位置
LOW とタグ付けされた修飾部
スコープに関してタグ付けされない ADVP 修飾部のディフォールト
LOWEST とタグ付けされた修飾部
述語
補部

引用文献

Allen, James F. (1983) “Maintaining Knowledge about Temporal Intervals.” *Communications of the ACM,* Vol. 26, No. 11, pp. 832-842.

Kamp, Hans, and Uwe Reyle (1993) *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language; Formal Logic and Discourse Representation Theory*, 2 Vols. Dordrecht: Kluwer Academic Publishers.

南不二男 (1974)『現代日本語の構造』大修館.

Pustejovsky, James (2017) “ISO-TimeML and the Annotation of Temporal Information.” In: N. Ide and J. Pustejovsky (eds.) *Handbook of Linguistic Annotation*. Dordrecht: Springer, pp. 941-968.

吉川克正・浅原正幸 (2015)「言語横断的手法による日本語時間的順序関係推定」，『言語処理学会第 21 回年次大会発表論文集』353-356.

５．主な発表論文等

〔雑誌論文〕　計12件（うち査読付論文　1件／うち国際共著　1件／うちオープンアクセス　0件）

１．著者名 Kei Yoshimoto	４．巻 -
２．論文標題 Modal Particles Yo and Ne in Japanese	５．発行年 2020年
３．雑誌名 Chungmin Lee and Jinho Park (eds.) Evidentials and Modals	６．最初と最後の頁 534-546
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 吉本啓・プラシャント-パルデシ	４．巻 2
２．論文標題 統語・意味解析情報付き日本語コーパスの構築	５．発行年 2020年
３．雑誌名 KLS Selected Papers 2: Selected Papers from the 44th Meeting of The Kansai Linguistic Society	６．最初と最後の頁 196-211
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

１．著者名 吉本啓	４．巻 -
２．論文標題 時間の言語的意味のコーパス化―日本語テンス・アスペクト表現理解過程解明に向けて	５．発行年 2021年
３．雑誌名 嶋田珠巳・鍛治広真『時間と言語』	６．最初と最後の頁 202-216
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 吉本啓・アラステア-バトラー・プラシャント-パルデシ	４．巻 -
２．論文標題 日本語ツリーバンクからの動詞格フレームの抽出	５．発行年 2021年
３．雑誌名 言語処理学会第27回年次大会発表論文集	６．最初と最後の頁 508-512
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 周振・吉本啓	4．巻 17
2．論文標題 統語・意味情報付きコーパスの開発に関する研究：中国語名詞句の解析について	5．発行年 2019年
3．雑誌名 国立国語研究所論集	6．最初と最後の頁 35-65
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 林則序・森 芳樹	4．巻 135
2．論文標題 懸念標識の構成的意味論に向けて	5．発行年 2019年
3．雑誌名 森 芳樹（編）『情報構造と話し手の状況把握』．日本独文学会叢書	6．最初と最後の頁 54-79
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 Katsumasa Ito and Yoshiki Mori	4．巻 -
2．論文標題 A Mirative Evidential in Exclamative	5．発行年 2019年
3．雑誌名 Proceedings of the 14th Workshop of Altaic Formal Linguistics (WAFL14). MITWPL	6．最初と最後の頁 117-128
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 Yuto Yamazaki and Yoshiki Mori	4．巻 1
2．論文標題 Kontrastivitaet und Wortstellung in deutschen Spaltsaetzen	5．発行年 2019年
3．雑誌名 Wie entsteht Bedeutung? Semantik zwischen Grammatik, Kognition und Kontext Iudicium Verlag	6．最初と最後の頁 73-88
掲載論文のDOI（デジタルオブジェクト識別子） 10.11282/jggls.1.0_73	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 Shungo Fujii and Yoshiki Mori	4．巻 2
2．論文標題 Verben, bei denen die lokalisierenden PPs eine Resultatslesart haben koennen	5．発行年 2020年
3．雑誌名 Japanische Gesellschaft fuer Germanistik (Hrsg.) Historische Syntax des Deutschen, Iudicium Verlag	6．最初と最後の頁 74-90
掲載論文のDOI（デジタルオブジェクト識別子） 10.11282/jggls.2.0_74	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 宮田瑞穂・森 芳樹	4．巻 -
2．論文標題 ドイツ語schonに基づいた日本語『もう』の分析	5．発行年 2020年
3．雑誌名 森 芳樹（編）『統語と意味のインターフェイスをめぐって―カートグラフィーの射程―』．日本独文学会叢書 NR.140	6．最初と最後の頁 58-72
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 Akari Takahata and Yoshiki Mori	4．巻 4
2．論文標題 Structure Removal in German Long Passive Constructions	5．発行年 2021年
3．雑誌名 Y. Ono & M. Shimada (Eds.) Data Science in Collaboration. Tsukuba: inext Co., Ltd	6．最初と最後の頁 58-67
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 Shungo Fujii and Yoshiki Mori	4．巻 4
2．論文標題 Experiencer-Argument im Haupt- und Komplementsatz der Einstellungsverben	5．発行年 2022年
3．雑誌名 Linguisten-Seminar: Forum japanisch-germanistischer Sprachforschung	6．最初と最後の頁 94-115
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計2件（うち招待講演　0件／うち国際学会　2件）

1．発表者名 周振・吉本啓
2．発表標題 “是 ... 的” 構造を取る中国語文の構造的曖昧性に関する考察
3．学会等名 The 21st Annual International Conference of the Japanese Society for Language Sciences（国際学会）
4．発表年 2019年

1．発表者名 Prashant Pardeshi, Kei Yoshimoto, Susanne Miyata, Koichi Takeuchi, and Hideki Kishimoto
2．発表標題 Development of a parsed corpus and its application to linguistic research and education
3．学会等名 The 21st Annual International Conference of the Japanese Society for Language Sciences（国際学会）
4．発表年 2019年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

—

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	森　芳樹 (Mori Yoshiki) (30306831)	東京大学・大学院総合文化研究科・教授 (12601)	

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

8．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関