

日本語研究のための日本語従属節意味分類基準の試案 —「鳥バンク」節間意味分類体系の再構築—

松本理美

立命館大学大学院文学研究科

gr0241xv@ed.ritsumei.ac.jp

1. はじめに

本稿では、日本語従属節¹へのアノテーションの精度を保つことを目的とした、日本語研究のための日本語従属節意味分類基準の試案を示す。節の意味分類において一定の基準を策定することにより、日本語従属節に関連した日本語研究の更なる進展が期待できる。意味分類基準は「鳥バンク」節間意味分類体系(池原, 2007)の再構築により策定する。『現代書き言葉均衡コーパス』(BCCWJ)(Maekawa et al. 2014) データ(一部)への従属節アノテーション²で明らかになった問題を基に、「鳥バンク」節間意味分類体系に修正を加えた日本語従属節意味分類基準の試案全容を概説し、特に補足節に着目した再編成³を詳細に検討する。

2. 「鳥バンク」節間意味分類体系の問題点

「鳥バンク」は、日本語-英語機械翻訳等の言語変換のために、日本語の重文と複文を対象として構築されたデータバンクである。「鳥バンク」節間意味分類体系は、「鳥バンク」の意味分類体系の一つで、主節と従属節の間の意味的な関係が4段階222項目に体系的に分類されている。従属節と主節を連結する600余りの「節間キーワード」と「日本文例」(以下、例文とする。)を各項目に配している。

松本・有田・浅原(2017)では、「鳥バンク」節間意味分類体系を基準としたBCCWJに対する従属節アノテーションの意味分類タグ付与において作業員間に生じた齟齬を分析し、従属節アノテーションの問題が節境界認定に関するものと、節の意味分類に関するものに分けられることを述べた。松本(2017)は、意味分類タグ付与の3次作業までの齟齬を再検討し、「鳥バンク」節間意味分類体系との関連を以下のように整理した。

- ① 「鳥バンク」節間意味分類による意味分類タグ付与作業の10%のタグに齟齬が生じている。
- ② タグ付与で生じた齟齬は「鳥バンク」意味分類体系に起因するものである。

- ③ 意味分類タグ付与の齟齬のうち、6割が副詞節と並列節の間に、3割が副詞節の下位分類に生じている。

①から「鳥バンク」節間意味分類体系が日本語研究を目的とした従属節の意味分類に有効であるが、転用することには若干の問題があると言える。②③から「鳥バンク」節間意味分類体系に起因する齟齬がいくつかの節に偏って生じていることがわかる。

そこで本稿では、「鳥バンク」節間意味分類体系に日本語学的視点で修正を加え、再構築することにより策定する日本語研究のための従属節意味分類基準の全容を概説する。また、3次作業において意味分類タグ付与の基準を設けたことにより、1, 2次作業の全ての齟齬が解消した補足節に限って、再編成の詳細を検討する。

3. 「鳥バンク」節間意味分類再構築の方針

「鳥バンク」節間意味分類は、日本語表現と英語表現が対応する設計となっており、日本語と英語の意味的対応を取るために、意味・形式による比較的細かい分類が行われている。しかし、この粒度の分類では、浅い段階から、意味分類の齟齬が生じる可能性が高く、アノテーションの精度が保てないという問題が起こり得る。

丸山他(2016)が述べている通り、日本語従属節の意味解釈には、文脈や書き手の意図によって変わるものが少なからずある。そのため、意味分類の自動化は極めて困難であり、現時点では人手に頼らざるを得ない。しかし、池原(2009)が述べているように、「重文・複文の持つ統語的な構造と意味の関係に着目すれば(中略)意味分類を半自動化することは可能であり、その意義は大きい。日本語従属節意味分類の半自動化にあたり、豊富な文型パターンからなる「鳥バンク」節間意味分類体系を基準とすることには、極めて大きなメリットがある。日英翻訳など言語の変換を目的として設計された「鳥バンク」節間意味分類体系を、日本語研究のための従属節意味分類基準に再構築するために、いかに修正するかが議論の中心となるところであろう。

¹ 本稿では、主節以外の節を全て従属節とする。節とは「複文を構成するところの、述語を中心とした各まとまり」(益岡・田窪, 1992)を指す。

² https://github.com/masayu-a/clause_pattern

³ 紙幅に限られているため、名詞修飾節、副詞節、並列節については稿を改める。

本稿では、従属節の意味・機能の分類には諸説あることを前提に、「鳥バンク」節間意味分類体系を再構築するための方針を以下の通り提案する。

- ① 日本語研究のための従属節の意味分類基準を策定することを目指し、基本的には益岡・田窪(1992)に則り、日本語学的な慣用に従う。
- ② 言語研究にとって貴重な研究成果である「鳥バンク」を発展させるという立場から、その修正は「鳥バンク」節間意味分類体系の転用により日本語従属節のアノテーションの精度が低下する可能性がある部分に着目して行う。

4. 「鳥バンク」節間意味分類体系再構築の全容

本節では、「鳥バンク」節間意味分類体系の再構築に

ついて、補足節以外の修正案を概説し、補足節に着目した修正案は次節で詳説する。

再構築における大きな修正点としては、「鳥バンク」節間意味分類体系の第1段階での補足節・名詞修飾節・副詞節・並列節の4分類を、補足節・連体節・連用節の3分類に修正する。分類項目数については、第2段階27項目を9項目に、第3段階37項目を20項目に、第4段階154項目を28項目に大幅に減じて再構成する。

「鳥バンク」節間意味分類体系の第2段階以降では、機能・意味・形式レベルでの分類の混在が見られる。修正案では第2段階を機能レベル、第3段階を意味レベル、第4段階を形式レベルでの分類とする。表1に「鳥バンク」節間意味分類体系の修正案の全容を示す。

表1 「鳥バンク」節間意味分類体系の修正案

統語	機能	意味	形式
第1段階	第2段階	第3段階	第4段階
補足節	名詞節		コト+格助詞等
			ノ+格助詞等
			トコロ+格助詞等
			節+格助詞等
	疑問節		選択疑問文
			疑問語疑問文
	引用節		直接引用
			間接引用・ト
			間接引用・ヨウニ
			間接引用・ナドト
連体節	補足語修飾節		間接引用・ナンテ
	内容節		同格関係
			非同格関係
	形式名詞修飾節		
	その他		

統語	機能	意味	形式
第1段階	第2段階	第3段階	第4段階
連用節	副詞節		時間節
			条件
			譲歩
			原因・理由・手段
			付帯状況
			様態
			逆接
			程度・比較
	目的		
	その他		
並列節	順接的並列		
	逆接的並列		

名詞修飾節については、分類名を連体節と改め、第2段階を6分類から4分類に減じる。第2段階の補足語修飾節下位の限定的・非限定的という分類を撤廃するなど第3段階以下では内容節の同格・非同格の分類だけに止め、19項目の詳細分類を5項目の粗い分類に修正する。

「鳥バンク」節間意味分類体系の第1段階における副詞節、並列節の分類を廃し、修正案ではこれらをまとめて連用節とする。第2段階で副詞節と並列節に分類し、下位分類は第3段階までとする。分類項目は可能な限り整理・集約し、180余りの細分類を12項目に止める。

詳細分類を避けるのは、意味分類タグ付与の齟齬の分析から、分類の粒度を細かくするほどアノテーションに揺れが生じることが明らかになったからである。アノテーターや作業によって節への意味分類タグ付与に齟齬が生じることは、データの信頼性やタグ付与の作業効率の低下を招く。「鳥バンク」節間意味分類体系を日本語研究に資するように再構築するためには、誰がいつアノテーションしても同じ結果が得られるような、精度の保たれた意味分類基

準として再編成しなければならない。

次に、補足節に着目した「鳥バンク」節間意味分類体系の再構築について、分類項目、「節間キーワード」や例文の再編を詳説する。

5. 補足節の再構築

松本(2017)では、BCCWJ データに対する従属節アノテーション作業において補足節に関連して生じた意味分類タグ付与齟齬の全てが「鳥バンク」節間意味分類体系の「節間キーワード」に起因したものであることを明らかにした。補足節は表層的な節末形式が標識となるため、節の意味分類に揺れが生じにくいと思われる。ところが、「鳥バンク」節間意味分類体系には、同形の「節間キーワード」が複数の節に存在するため、「節間キーワード」が意味分類タグ付与齟齬の原因となった。

そこで、意味分類タグ付与の齟齬の原因となった「節間キーワード」に着目し、分類項目の整理・集約を行った。「鳥バンク」節間意味分類体系では、補足節を下位の第2

段階で3項目に、第3段階で8項目に、第4段階で24項目に細分類している。本稿の試案では第2段階を機能レベルの、第3段階を意味レベルの、第4段階を形式レベルの分類とするため、第3段階では下位の分類を行わ

ず、第4段階で形式により11項目に下位分類する。表2に「鳥バンク」節間意味分類体系の補足節(左)と修正案(右)を併記して示す。

表2 「鳥バンク」節間意味分類体系(補足節)の修正案

「鳥バンク」節間意味分類体系				
第1段階	第2段階	第3段階	第4段階	
補足節(補語相当節)				
	名詞節	コト型	形式・意味により8項目に細分類	
		ノ型	強調構文	
		トコロ型	形式・意味により3項目に細分類	
		節+格助詞型		
	疑問節	選択疑問文		
		疑問語疑問文		
	引用節	直接引用		
		間接引用	ト型	
			トハ型	
			ヨウニ型	
			ナドト型	
			ナンテ型	
			選択疑問節型	
疑問語疑問節型				

「鳥バンク」節間意味分類体系の修正案			
第1段階	第2段階	第3段階	第4段階
補足節			
	名詞節		コト+格助詞等
			ノ+格助詞等
			トコロ+格助詞等
			節+格助詞等
	疑問節		選択疑問文
			疑問語疑問文
	引用節		直接引用
			間接引用・ト
			間接引用・ヨウニ
			間接引用・ナド
			間接引用・ナンテ

以下に、「節間キーワード」が意味分類タグ付与の齟齬の原因となることを、「鳥バンク」節間意味分類体系の例文を挙げて説明する。「節間キーワード」に下線を付し、この例文が分類されている項目を《第1段階、第2段階、第3段階、第4段階、》と記した。

(1) 丁度夕食が終わったところに彼がやってきた。《補足節、名詞節、トコロ型》

(2) 私が車を降りようとしているところにちょうど彼女がふらっと通りかかった。《副詞節、時、事態の時、進行》

(1)の「ところに」は補足節の「節間キーワード」であり、(2)の「ところに」は副詞節の「節間キーワード」となっている。ところが、日本語学的には(1),(2)の「ところに」に導かれる節はともに形式名詞「トコロ」が格助詞「に」を伴った副詞節であるとも、時を表す副詞節であるとも解釈できる。

(3) 僕の見たところでは彼が一番だ。《副詞節、場面、視点》

(3)の「ところでは」は、副詞節の「節間キーワード」となっているが、日本語学的には「ところ」に格助詞「で」と提題助詞「は」が後接している補足節であるとも解釈できる。

同形の「節間キーワード」が注記なしに複数の節の「節

間キーワード」として存在しているのは望ましくない。日本語従属節には例文(1)(2)のように複数の解釈が可能である節や意義同形の節末形式を持つ節も少なからず存在する。また、「鳥バンク」節間意味分類体系では、「ことだ」「のだ」「ところだ」を補足節の「節間キーワード」としているが、この分類の是非にも日本語学上の諸説がある。このように判定の難しいケースはあるが、アノテーションの精度を保つためには一定の分類基準や枠を定める必要がある。

そこで、補足節下位の名詞節節末形式についての規定を以下の通り提案する。なお、「鳥バンク」節間意味分類体系の「節間キーワード」を日本語学の慣用に寄せ、以降は節末形式と呼ぶことにする。

- I. 格助詞等を伴う「こと」、「の」、「ところ」は全て補足節の節末形式とする。
- II. 格助詞等とは、格助詞、提題助詞、取り立て助詞と格助詞相当(益岡・田窪, 1992:49-50, 81)および複合辞・助詞相当句の格助詞⁴(小椋他, 2011:資料 2-9)である。(表3参照)
- III. テンス・アスペクトを表す述語に伴う「ところ(に、へ)」が副詞節の節末形式ともなり得るなど、複数の意味が認められる場合は、同形の節末形式が他の項目にもあることを注記して、参照できるようにする。
- IV. 異義同形の節末形式が他の項目にある場合も、その旨を注記し、参照できるようにする。

⁴ 形式名詞「コト」「ノ」「トコロ」に後接することができない複合辞、連体形複合辞を除く。

- V. 文末表現としての「ことだ」「のだ」「ところだ」については、補足節下位の名詞節に分類しない。
- VI. 第 4 段階では、「選択疑問節」「疑問語疑問節」による分類を行わない。

表3 格助詞等

格助詞	提題助詞	取り立て助詞	複合辞 (格助詞相当)	
が	は	は	からして	につき
を	なら	も	からすると	にとつて
に	って	さえ	からすれば	によつて
から	ったら	でも	として	により
と		すら	にあたって	によると
で		だって	にあたり	によれば
へ		まで (に)	に至るまで	にむかつて
まで		だけ	において	に沿つて
より		ばかり	にかけて	につれて
		のみ	に関して	にしたがつて
		しか	に際し	を通じて
		こそ	に際して	をはじめ
		など	にして	を通して
		なんか	に對し	をめぐつて
		なんて	に對して	をともなつて
		くらい	について	を介して

上記の規定に則って「鳥バンク」節間意味分類体系を見直すと、例えば、格助詞相当の複合辞が追加されたことにより、副詞節に分類されていた「節間キーワード」と例文が補足節に再編される。以下にその例を示す。

- (4) 聞くところによると、今度新しい国語辞典が出るそうだ。《副詞節, 場面, -, 情報源》
- (5) 私が過去に調査したところによれば、その結果は正しい。《副詞節, 場面, -, 情報源》

(4)(5)は、いずれも副詞節に分類されており、この分類項目はこの 2 例のみからなっている。格助詞相当の複合辞に関する規定を追加したことで、節末形式「ところによると」「ところによれば」は補足語下位の「トコロ + 格助詞等」の節末形式となり、(4)(5)は補足節に分類されることになる。節末形式が意味分類の標識として機能し、副詞節の下位分類が減少することによって、揺れない意味分類アノテーションが可能になる。このような例は他に 22 例あり、この規定に則って整理・集約することにより副詞節の下位の 8 項目が不要となる。

補足節下位の引用節では、間接引用の下位 (第 4 段階) に設けている選択疑問節型と疑問語疑問節型の分類を撤廃することで、「節間キーワード」の「か」と「かしら」となど引用の「と」を伴う節は全て間接引用のト型に集約できる。

以上のように、規定に則って分類項目を整理・集約することで、補足節の第 4 段階 5 項目、副詞節の第 4 段階 9 項目を撤廃することが出来る。

6. おわりに

本稿では、「鳥バンク」節間意味分類体系の再構築による日本語従属節意味分類基準の試案全容と補足節における修正の詳細について検討した。日本語学においては、形式名詞「ノ」を含む文末表現「のだ」だけでも膨大な研究があり、その議論の歴史も長い。議論の決着を待ってはいは、いつまでも日本語従属節意味分類基準の策定ができない。佐藤他 (2016) の言葉を借りるなら、「ある方針に従って決めるしかない」ということである。本稿に示した試案は、今後の議論の基になるものであり、節のアノテーション作業などを通して、改良を重ねるべきものである。

意味分類基準の策定に加え、意味分類の揺れがあることを許容し、1 つの節に複数のアノテーションを認めるなど、精度を保つためのアノテーション基準の策定も今後の課題として挙げられる。また、節境界の認定基準の策定というさらに悩ましく、大きな課題も避けては通れない。

謝辞

本研究 (の一部) は国立国語研究所コーパス開発センターの共同研究プロジェクト「コーパスアノテーションの拡張・統合・自動化に関する基礎研究」(2016-2021 年度) の成果である。

参考文献

- 池原悟 (2007) 「意味類型パターン記述言語仕様書」
- 池原悟 (2009) 『非線形言語モデルによる自然言語処理』岩波書店
- Maekawa Kikuo, Yamazaki Makoto, Ogiso Toshinobu, Maruyama Takehiko, Ogura Hideki, Kashino Wakako, Koiso Hanae, Yamaguchi Masaya, Tanaka Makiro, and Den Yasuharu (2014) Balanced Corpus of Contemporary Written Japanese, *Language Resources and Evaluation (LRE48)*:345-371
- 益岡隆志・田窪行則 (1992) 『基礎日本語文法』くろしお出版
- 松本理美・浅原正幸・有田節子 (2017) 『現代日本語書き言葉均衡コーパス』に対する節の意味分類情報アノテーション - 基準策定、仕様書作成の必要性について - 『言語資源活用ワークショップ 2016 発表論文集』336-346
- 松本理美 (2017) 「従属節の意味分類基準策定について - 鳥バンク基準互換再構築の検討 -」 『言語資源活用ワークショップ 2017 発表論文集』39-50
- 丸山岳彦・佐藤理史・夏目和子 (2016) 「現代日本語における節の分類体系について」 『言語処理学会第 22 回年次大会発表論文集』1113-1116
- 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕 (2011) 『現代日本語書き言葉均衡コーパス』形態論情報規定集第 4 版 (上) 特定領域研究「日本語コーパス」平成 22 年度研究成果報告書
- 佐藤理史・丸山岳彦・夏目和子 (2016) 「現代日本語書き言葉均衡コーパスに対する節境界付与」 『言語処理学会第 22 回年次大会発表論文集』409-412