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　元 年　 ６ 月 １８ 日現在

機関番号：１２４０１
研究種目：若手研究(B)
研究期間：2015～2018
課題番号：１５Ｋ１６７６４

研究課題名（和文）統語・意味情報付き形容詞を実装した通時コーパスによる中古形容詞の意味・用法研究

研究課題名（英文）Construction and Analysis of Adjective's Syntactic and Semantic Annotation of the 'Corpus of Historical Japanese' Heian Period Series

研究代表者
池上　尚（IKEGAMI, Nao）

埼玉大学・教育学部・准教授

研究者番号：５０７３９１２５

交付決定額（研究期間全体）：（直接経費）　　3,100,000円

研究成果の概要（和文）：日本語史研究の新たな言語資源「日本語歴史コーパス」に、従来実装が十分でなかった統語・意味情報をアノテーションし、これを活用した中古形容詞の意味・用法について研究を進めた。統計的手法を用いて、客観性の担保された複合形容詞の語認定を行ったほか、当代の複合形容詞が文にパラフレーズ可能な語性を有することを計量的に示した。

研究成果の学術的意義や社会的意義
既刊の総索引では不十分であった、計量的分析に基づく語認定に取り組み、対象となる語の周辺に位置する語的な表現にまで視野を広げた中古形容詞の実態解明に努めた。また、古典語の語構成は語彙論・意味論の枠組みにおいて分析されるのが主流であったが、複合形容詞における語と文との連続性を計量化し、文法論と関与する語構成研究の一例を示した。

研究成果の概要（英文）：This project created of syntactic and semantic annotation for 'Corpus of Historical Japanese' Heian Period Series. In order to refine semantic studies of compound adjectives of Heian period era, using Quantitatibe Analysis.

研究分野：日本語学

キーワード：中古語　形容詞　コーパス　語彙論　意味論　語構成論

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９、ＣＫ－１９（共通）

１．研究開始当初の背景

国立国語研究所が公開している「日本語歴史コーパス」は、日本語史研究の新たな言語資源として注目を集めている。現在公開されているコーパスには、品詞や活用型、原文の文字列といった形態に関する情報が実装されており、これを利用した日本語史研究は盛んに進められている。しかし、語どうしの主述関係といった統語情報や、文脈レベルの意味情報といったものは十分に実装されていない。文法史や語彙史の分野においてコーパスを用いた計量的な研究を進めていくにあたり、こうした新たなアノテーションが必要である。

２．研究の目的

本研究では、(A)既存の「日本語歴史コーパス 平安時代編」（以下、「CHJ 平安編」と呼ぶ）における形容詞を中心に、これまで十分でなかった統語・意味情報をアノテーションし、コーパスの一層の充実を図るとともに、(B)これを活用した中古形容詞の意味・用法の研究を行う。

(A)については、「CHJ 平安編」ですでに付与されている形態情報に加え、新たに、形容詞の“「主体」（＝主語・被連体修飾語・被連用修飾語）という統語情報”、“見出し語レベルの意味分野情報／文脈レベルでの意味情報”をアノテーションする。

(B)では、(A)によって構築される形態・統語・意味情報付き形容詞を実装したコーパスを活用し、中古形容詞の部分体系（数語の形容詞どうしの関係）・総合的な体系（部分体系どうしの関係）を体系づけている要素を分析する。

３．研究の方法

(1)　「CHJ 平安編」に出現する形容詞（異なり約 640 語、延べ約 38000 語）に、宮島達夫他編(2014)『日本古典対照分類語彙表』（笠間書院）で示された見出し語レベルの意味分野情報である番号を振る。

(2)　(1)で付与した番号により形容詞をグルーピングし、意味分野別に作業単位の分割を行う。また、作業順を決定する。

(3)　作業単位ごとに人手で形容詞の主体を抽出し、統語情報、文脈レベルの意味情報をタグ付けする。

(4)　(3)の作業が済んだデータから統計的手法を用いて数量化し、中古形容詞を体系づける要素の分析を行う。

４．研究成果

(1)　分類語彙表番号「31332(良不良・適不適)」をもつ形容詞のうち、「CHJ 平安編」、「日本語歴史コーパス 鎌倉時代編Ⅰ説話・随筆」（以下、「CHJ 鎌倉編Ⅰ」と呼ぶ）に出現する評価形容詞４語「良し」「悪（あ）し」「悪（わろ）し」「悪（わる）し」、「良し」の類義語である「宜し」（分類語彙表番号「31331(特徴)」）を加えた計５語に対し、主体たる名詞の同定、および、主体・形容詞の構文環境（主述関係における助詞・副詞等の有無）を調査し、構文バリエーションの整理を行った（表１）。

表1　中古･中世前期における名詞―評価形容詞の構文バリエーション

類		構文							
Ⅰ	**A**				名詞				形容詞
	A+			程度副詞					
Ⅱ	**a-1**		接頭辞「御」						
	a-2	連体句							
	B				名詞	助詞			形容詞
	B+			程度副詞					
	b-1		接頭辞「御」						
	b-2	連体句							
	C				名詞	助詞	助詞		形容詞
	C+			程度副詞					
	c-1		接頭辞「御」						
	c-2	連体句							
	D				名詞			副詞	形容詞
	d-1		接頭辞「御」						
	d-2	連体句							
	E				名詞	助詞		副詞	形容詞
	e-1		接頭辞「御」						
	e-2	連体句							
	F				名詞	助詞	助詞	副詞	形容詞
	f-1		接頭辞「御」						

太字:「CHJ平安編」にのみ出現

中古では 20 の類が認められるが、中世前期には 16 の類となり、中古における複合形容詞（的表現）の構文バリエーションの複雑さがうかがえる。従来の総索引や語彙表ではこうした構文バリエーションについて知ることができなかったが、本研究では既存の「CHJ 平安編」「CHJ 鎌倉編Ⅰ」に新たな統語情報をアノテーションすることで、複合語とその周

辺に存在する文との連続性を視野に入れた語構成研究が可能になりつつある。

次に、このデータのうち、主体たる名詞と評価形容詞とが直接続くⅠ類（A・A+）について、語認定の問題に取り組んだ。そもそも助詞が必ずしも明示的でない古典語においては、どこから１語と認めるべきか判断が難しく、語認定における客観性の担保が課題となる。そこで、名詞・評価形容詞のコロケーション強度を検出し、１語として認定すべき［名詞＋評価形容詞］の候補を定めた。統計的指標はｔスコア（高頻度・一般性の高いコロケーション検出に有効）・ＭＩスコア（低頻度ながら特殊な結びつきをしているコロケーション検出に有効）の２種を用い、それぞれ有意とされる数値、ｔスコア≦2.0、ＭＩスコア≦3.0 の場合に１語と認定した。中古・中世前期に共通しているのは、Ⅰ類のうち、コロケーション強度の強い１語と認められる複合形容詞であっても、Ⅱ類に展開しうるという点である。すなわち、意味変化を生じずに NA 形式と NpA 形式とを行き来する、文にパラフレーズ可能な複合形容詞であることが分かった。現代語においては、こうした「複合形容詞と文との近接現象」（山本清隆 1996）が指摘されることはあったが、古典語における実態を計量的に示した研究はない。名詞と評価形容詞の組み合わせという限定された範囲ではあるが、本研究によって、中古から中世前期における複合形容詞の語としての特異性が明らかになった。【５．〔雑誌論文〕③、〔学会発表〕③】

(2)　「CHJ 平安編」所収の『竹取物語』『土佐日記』２作品に出現する語すべてについて、分類語彙表番号アノテーションの試行をし、類（体・用・相・他）の比率（表２）、部門（関係・主体・活動・生産物・自然・枕詞）の比率（表３）、上位５つの分類項目を分析した。統語的（類）には２作品に差が見られるが、意味的（部門）にはあまり差が見られないという結果が得られた。【５．〔学会発表〕①】

表２　類

	1 Noun 体	2 Verb 用	3 Adj, Adv 相	4 Other 他	Total
Taketori 竹取物語	2318 43.3%	2252 42.1%	706 13.2%	72 1.3%	5348
Tosa 土佐日記	1710 49.1%	1272 36.5%	454 13.0%	45 1.3%	3481
All	4028 45.6%	3524 39.9%	1160 13.1%	117 1.3%	8829

表３　部門

	0.1 Relation 関係	0.2 Subject 主体	0.3 Action 活動	0.4 Product 生産物	0.5 Nature 自然	0.9 Poetic 枕詞	Total
Taketori 竹取物語	2335 43.7%	577 10.8%	1722 32.2%	229 4.3%	485 9.1%		5348
Tosa 土佐日記	1540 44.2%	360 10.3%	1002 28.8%	167 4.8%	411 11.8%	1 0.0%	3481
Total	3875 43.9%	937 10.6%	2724 30.9%	396 4.5%	896 10.1%	1 0.0%	8829

また、特に、相の類の中で自然部門の番号「3.50」をもち、かつ、複数の番号をもつという曖昧性が問題になる形容詞 22 語（延べ語数 1447 語）に対し、コーパスの底本である新編日本古典文学全集（小学館）での現代語訳、当該形容詞の表す属性が帰属する原文での主体、文脈レベルでの分類語彙表番号、以上３種の情報をアノテーションした。この作業を経て、今後解消すべき課題として、連語や慣用表現中の単位へのアノテーションがあることなどを指摘した。【５．〔学会発表〕②】

(3)　「CHJ 平安編」「CHJ 鎌倉編Ⅰ」に頻出し、動作・状態の程度が大きいことを強調する語りの定型表現「～事限り無し」について、上接語を抽出し、コーパスの底本である新編日本古典文学全集（小学館）において「～事限り無し」全体に当てられている現代語訳を新たな意味情報としてアノテーションした。『今昔物語集』に出現する「～事限り無し」は動詞を上接しやすいという指摘がすでになされているが、今回の調査により、中古から中世前期を通じ、そうした傾向が共時的に認められるということも明らかになった（表４）。

表４　「事限り無し」の上接語

品詞(大分類)	「CHJ平安編」		「CHJ鎌倉編Ⅰ」	
	頻度	比率	頻度	比率
形容詞	32	28.3%	138	23.8%
形状詞	6	5.3%	17	2.9%
名詞	0	0.0%	1*	0.2%
動詞	75	66.4%	422	72.9%
解釈不明	0	0.0%	1*	0.2%
総計	113	100.0%	579	100.0%

*誤解析と推測される例

また、定型表現ゆえに特定の現代語には置き換えにくく、上接する名詞句を含めた「～事限り無し」全体を、文脈に沿って他の多様な表現に置き換えなければ自然な現代語訳とならない点に着目し、意訳を目的とする古文教材のデータにも展開した。さらに、中学校教員との連携を図り、このデータを用いた授業を実際に行い、教材作成からその効果についても検証した。【５．〔雑誌論文〕②、〔図書〕①】

(4)　分類語彙表番号「3.1911(長短・高低・深浅・厚薄・遠近)」をもつ形容詞のうち、名詞を前項とする複合形容詞としての用法（名詞との熟合度の強弱）が問題になる「長し」「短し」「高し」「低し」「深し」「浅し」「厚し」「薄し」を対象に、「CHJ 平安編」からデータを抽出し、主体たる名詞の同定、および、主体・形容詞の構文環境（主述関係における助詞・副詞等の有無）を調査した。構文バリエーションの整理は名詞―評価形容詞（表１）と比較する形で行った（表５）。

表5　中古における名詞―次元形容詞の構文バリエーション

類		構文							
Ⅰ	**A**			程度副詞	名詞				形容詞
	A+								
Ⅱ	**a-1**		接頭辞「御」						
	a-2	連体句							
	a-3	連体句	接頭辞「御」						
	B			程度副詞	名詞	助詞			形容詞
	B+								
	b-1		接頭辞「御」						
	b-2	連体句							
	b-3	連体句	接頭辞「御」						
	C			程度副詞	名詞	助詞	助詞		形容詞
	C+								
	c-1		接頭辞「御」						
	c-2	連体句							
	D				名詞			副詞	形容詞
	d-1		接頭辞「御」						
	d-2	連体句							
	E				名詞	助詞		副詞	形容詞
	e-1		接頭辞「御」						
	e-2	連体句							
	e-3	連体句	接頭辞「御」						
	F				名詞	助詞	助詞	副詞	形容詞
	f-1		接頭辞「御」						

太字：名詞―次元形容詞として実際に出現

中古における名詞―次元形容詞には 17 の類が認められ、一見、同時代の名詞―評価形容詞の構文バリエーションより単純化しているように思われる。しかし、中古・中世前期における名詞―評価形容詞になかった新たな類（a-3、e-3。ともに連体句と接頭辞「御」を伴う）が認められた。こうした構文バリエーションの差異は、後項形容詞の意味分野に起因するものであるのか、あるいは、前項名詞が関与する問題であるのか、今後の課題とする。

次に、このデータのうち、主体たる名詞と評価形容詞とが直接続くⅠ類（A・A+）について、語認定の問題に取り組んだ。コロケーション強度の検出方法は研究成果(1)と同様である。分析の結果、１語として認定すべき［名詞＋次元形容詞］の約６割が、コロケーション強度の強いⅠ類として熟合しながらⅡ類に展開しうる語性をもち、［名詞＋評価形容詞］と共通することが明らかになった。評価形容詞や次元形容詞といった特定の意味分野ではあるものの、古典語において語と文とを横断する表現の実態を計量的に示した研究はこれまでにない。今後は、他の意味分野の形容詞も含めた体系的な分析を目指すことで、語構成論と文論とを繋ぐ意味論研究を進めていく。【５．〔雑誌論文〕①】

(5)　「3.1911(長短・高低・深浅・厚薄・遠近)」「31400(力)」「31910(多少)」「31912(広狭・大小)」「31913(速度)」「33012(恐れ・怒り・悔しさ)」の分類語彙表番号をもつ形容詞のうち、複合形容詞としての用法（名詞との熟合度の強弱）が問題になる一部の語について、「CHJ 平安編」からデータを抽出し、主体の同定、および、主体・形容詞の構文環境（主体―形容詞の主述関係における助詞・副詞等の有無）を調査した。今後は、研究成果(1)(4)の分析手法の妥当性を検討しつつ、データの数量化を行う。

(6)　「CHJ 平安編」「CHJ 鎌倉編Ⅰ」に出現する〈言いようもない〉を表す慣用表現「言はむ方なし」をはじめとして、発話を表す動詞を含む表現を抽出し、コーパスの底本である新編日本古典文学全集（小学館）での現代語訳、当該表現の表す属性が帰属する原文での主体、統語環境、以上３種の情報を付与した。今後は、研究成果(3)の「～事限り無し」のデータと比較しながら、高程度を強調する表現の体系について意味記述を進める。

(7)　「CHJ 平安編」に出現する高程度を表す形容詞「いみじ」（延べ 1754 語）に対し、コーパスの底本である新編日本古典文学全集（小学館）での現代語訳、当該形容詞の表す属性が帰属する原文での主体、統語環境、以上３種の情報を付与した。今後は、古文教材のデータとしての応用方法を検討し、授業実践とその効果について検証する。

＜引用文献＞
①山本清隆、複合語と文の境界、日本語学、15-9、1996、pp.41-49

５．主な発表論文等

〔雑誌論文〕（計３件）

① 池上尚、中古語複合形容詞［名詞＋次元形容詞］の構文バリエーション、日本語文法史研究、査読無、4、2018、pp.23-42
② 池上尚、甲斐伊織、高程度表現「～事限り無し」に着目した授業実践の試み―『日本語歴史コーパス』を活用して―、埼玉大学国語教育論叢、査読有、20、2017、pp.43-54
③ 池上尚、中古語複合形容詞［名詞＋評価形容詞］の一語性、国語語彙史の研究、査読有、35、2016、pp.39-55

〔学会発表〕（計３件）

① Masayuki Asahara, Nao Ikegami, Yutaka Hara, Sachi Kato and Tai Suzuki、Annotation of ‘Word List by Semantic Principles’ Labels for ‘Corpus of Historical Japanese’ Heian Period Series -- Trial Annotation on Tosa Nikki and Taketori Monogatari --、Japanese Association for Digital Humanities2017、2017
② 池上尚、『日本語歴史コーパス 平安時代編』出現形容詞に対する古典分類語彙表番号アノテーション、言語処理学会第 23 回年次大会、2017
③ 池上尚、コーパスに見る中世語複合形容詞の一語性、「通時コーパス」国際シンポジウム、2015

〔図書〕（計１件）

① 河内昭浩、池上尚、小木曽智信、小林正行、杉山俊一郎、鈴木泰、須永哲矢、服部紀子、宮城信、渡辺由貴、甲斐伊織、下田俊彦、朝倉書店、新しい古典・言語文化の授業―コーパスを活用した実践と研究―、
2019、pp.34-46、pp.59-71

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。