

『虎明本狂言集』のコーパスデータにおける短単位認定の諸問題

渡辺由貴・市村太郎・鴻野知暁 (国立国語研究所コーパス開発センター)

Problems Concerning the Recognition of Short-Unit-Word in the Toraakira-bon Kyogensyuu Corpus

Yuki Watanabe Taro Ichimura Tomoaki Kouno

(National Institute for Japanese Language and Linguistics)

要旨

『虎明本狂言集』のコーパスデータの作成・整備過程で、語（短単位）の認定を行う必要があるが、その際に困難が生じる場合がある。例えば、仮名で表記された同音の語の認定、活用語尾が表記されていない語の音便形の認定、形容詞連用形の文語活用・口語活用の認定等である。

同音の語については、底本の用例の状況や校注者の傍記を元に認定を行った。音便形の認定については、四段活用動詞のうち連用形の用例数の多い語について後接語別の音便状況の調査を行い、明らかに読みが予想できる例以外については、「た」が後接するもののみを音便形とし、それ以外の語が後接するものについては無理に音便形を認めない方針とした。形容詞の活用の認定については、形容詞の終止形活用語尾と連体形活用語尾の状況を調査した結果、形容詞の口語活用化が進んでいたと判断し、明らかな文語活用の例を除き、口語活用を原則とした。

1. はじめに

国立国語研究所『日本語歴史コーパス』構築の一環として進められている『虎明本狂言集』のコーパスデータの作成・整備過程で、語（短単位）の認定を行う必要があるが、1642年に成立した『虎明本狂言集』は、その言語事象が古代語から近代語・現代語への過渡の特徴を示しており、語の認定において困難が生じる場合がある。

例えば、「異見一意見」「時宜一辞儀」のように、類似した二つ以上の語が『虎明本狂言集』の成立時期に混在し、時に混同され用いられていることがあるが、このような語についても、コーパスデータ上は、いずれかの語と認定することが必須となる。

また、活用語の音便形についても類似した問題がある。例えば、『虎明本狂言集』において、「いたいてーいたして」のように音便形と非音便形の両表記形がみられる語があるが、「致て」のような活用語尾の表記されていない語形があらわれた場合、その活用形を音便形か非音便形かのいずれかに認定する必要がある（市村 2014, pp.106-107）。

形容詞の活用型についても、『虎明本狂言集』においては文語特有の活用語尾「ーし」「ーき」と口語特有の活用語尾「ーい」の両形が見られ、例えば形容詞「長い」の連用形「ながく」を、文語活用か口語活用かのいずれかに認定しなくてはならない。

これらの問題については、底本の注釈や索引、各種辞書の記述、研究論文等が参考になるが、これらを参照しても『虎明本狂言集』におけるそれぞれの語を確定するには至らないこともある。例えば、注釈や索引において、“二つの語のどちらの可能性もある”という示し方がされている場合があり、これは実態に即した記述であるが、コーパスデータにおいてはそのような曖昧な処理はできない。さらに、本コーパスは、『日本語歴史コーパス』の中の一つのコーパスであるため、中古語から近代語、現代語のコーパスで蓄積され

たデータの中に位置づける必要がある。

本発表では、『虎明本狂言集』のコーパスデータにおいて語の認定が難しい事例をとりあげ、注釈や索引、辞書等を参照しながら検討したい。

2. 意味・用法の類似する同音の語の認定

『虎明本狂言集』においてみられる、意味・用法が類似する同音の語の認定について検討するにあたり、まず、国立国語研究所のコーパスデータにおける同語異語判別の方針を確認しておく。コーパスデータは自動形態素解析を前提としており、その精度を保つために、コーパスデータにおける同語異語の判別については次のような方針が立てられている。

方針1：同表記異語を生じさせるような語彙素の立て方はできる限り行わない。

方針2：複数の語彙素に分ける場合は、明確な基準・理由をもってし、人手で正確に区別できないような語彙素の分割は行わない。(小椋他 2011, p.137)

現代語のコーパスデータにおいては、BCCWJ から取得した頻度情報や、『岩波国語辞典』第6版、『国語大辞典』、『大辞林』『広辞苑』における見出しの立て方等を考慮しながら同語異語判別を行っている。例えば、動詞「アウ」については「合う」「会う」の二つの語彙素を立て、「逢う・遭う・遇う」は「会う」の書字形としている。動詞「オサマル」については、「収まる」のみを語彙素として立て、「治まる」「納まる」「修まる」等は全てその書字形としている(小椋他 2011, pp.137-140)。

しかし、中古・近代および現代語のコーパスを作成する過程で、別の語彙素として登録されている二語が、『虎明本狂言集』においては明確に別語であると判別できない場合が存在する。資料の成立時期に、類似した二つ以上の語が混在し、時に混同され用いられているケースがあること、表記にゆれがあること、現代と中近世とでその語の意味や表記が異なっているケースがあること等がその理由で、上記の基準では語の判別・認定に迷うことがある。例えば、「意見—異見」、「憂き世—浮き世」、「辞儀—時宜」、「卑怯—比興」等がそれにあたる。以下、「意見—異見」を例に見ていく。

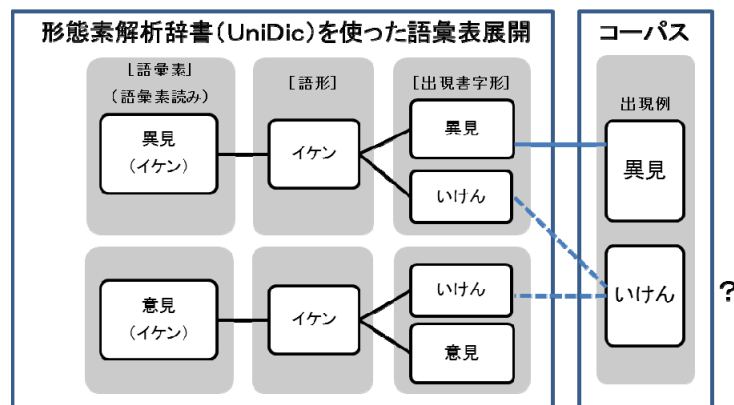


図1 同音異義語の認定 「意見—異見」

『虎明本狂言集』において見られる「語彙素読み」が「イケン」の語は、既登録の「意見」「異見」のいずれかに相当するものである。現代語において、「異見」は、特に他者

と異なった見解である場合に用い、また、その意味の場合のみ「異見」の表記をするのが一般的であると考えられるため¹、「意見」であるか「異見」であるかは意味・表記にしたがって判別することが可能であろう。また、古くは「意見」が「本来は政務などに関する衆議の場において各人が提出する考えであった」（『日本国語大辞典 第二版』）ことから、「意見」と「異見」とは別語と認識されていたようである。しかし、中世頃にはこの両語は混同されて用いられることがあったようであり、明確に区別することが難しく、慎重な判断が必要となる。

『虎明本狂言集』における「イケン」の表記別の用例数は、「異見」表記 7 例、「いけん」表記 11 例である。「方針 1：同表記異語を生じさせるような語彙素の立て方はできる限り行わない。」により、「異見」表記の 7 例については「意見」としない方が望ましく、「異見」とするのが妥当であると考えられるが、仮名表記の「いけん」については、漢字表記例に合わせて一律「異見」とするべきか、あるいは「意見」の可能性があるのか、検討が必要になる。

『日本国語大辞典 第二版』および『時代別国語大辞典』は、「異見」と「意見」を同一見出しの中に立てている²。また、『日本国語大辞典』の「語誌」によると、「意見」と「異見」は明治になると典拠主義の辞書編纂の立場から別の語とされるようになるが、中世後期の古辞書類や文学作品では「異見」が一般的であったとされており、『時代別国語大辞典』でも「次第に『異見』が『意見』の領域を侵して、両者の区別が失われがちであった」との記述がある。

また、『大蔵虎明本狂言集総索引』の各巻でも、「イケン」の語の見出し字が異なり、大名狂言、女狂言、萬集類の索引が「意見」（「御意見」「御意見有る」を含む）としている一方、鬼・小名類、出家座頭類、集類の索引は「異見」（「御異見」を含む）としている。また、聳・山伏類の索引については、「ごいけん [御異見・御意見]」と、両方の表記を見出し字としている。「異見」「意見」の両語は区別しがたいものであり、結果的にその巻の担当者の判断によって見出し字が分かれることになったと想像される。

用例を確認すると、「異見」表記の例、仮名表記「いけん」の例とも、「忠告」「助言」の意味と解釈可能な例である。これらの例では、「他者と異なる見解」といったニュアンスは強くなく、現代語であれば「意見」と表記するのが一般的に思われるような例ではある。ただし、『虎明本狂言集』には漢字表記「意見」の例があらわれず、底本の校注者も仮名表記「いけん」7 例のうち 6 例に「異見」と傍記している。さらに、次の例のように、「異見」表記の例と「いけん」表記の例との間に意味の違いは認めがたく、「異見」表記の例を語彙素「異見」とするのであれば、仮名表記「いけん」の例も語彙素「異見」とするのが妥当と考えられる。

- (1) (新座の者) 今日よりは、かた / \ をよりおや殿とたのみまらす程に、万事よひやうに引まはされて、御【いけん】有てくだされい（鼻取りずまふ 上 p.196³）

¹ 例えば、『岩波国語辞典』（第六版）では、以下のように立項されている。

【意見】①ある問題についての考え。②自分の考えを述べて人をいましめること。

【異見】他と違った考え。

² ただし、『日本国語大辞典』は「意見・異見」の見出しと別に「異見」も立項している。

³ 以下、引用は大塚（2006）による。

- (2) (親) しつけもなひやつで御ざる程に、今からは萬事御【異見】たのみまらす
(二人袴 上 p.417)
- (3) (伯藏主=狐) かやうにいふて又つたと云事をきひたらば、二たびてらへもなおりや
つそ、その【いけん】いたさうとぞんじて参た
(男) 近比かたじけなふ御ざる、私をおぼしめせばこそ、さやうの御【異見】を
なされてくださるれ (つりきつね 下 pp.418-419)

このように、用例や、校注者の傍記等の状況から、『虎明本狂言集』における「イケン」は全て「異見」と判断するのが妥当であろう。

3. 音便形の認定

活用語の音便形の認定についても困難な例がある。市村 (2014, pp.106-107) にあるように、『虎明本狂言集』においては「いたいて—いたして」のように音便形と非音便形の両表記形がみられる語があり⁴、「致て」のような活用語尾の表記されていない語形があらわれた場合、その活用形を音便形である「連用形 - イ音便」とすべきか、非音便形「連用形 - 一般」とすべきかについて判断する必要がある。『日本語歴史コーパス』のうち、中古和文のデータにおいては、基本的に活用語尾が仮名表記されているためにこのような問題は起こりにくいと考えられ、これも中世語資料のデータゆえの問題であると言える。

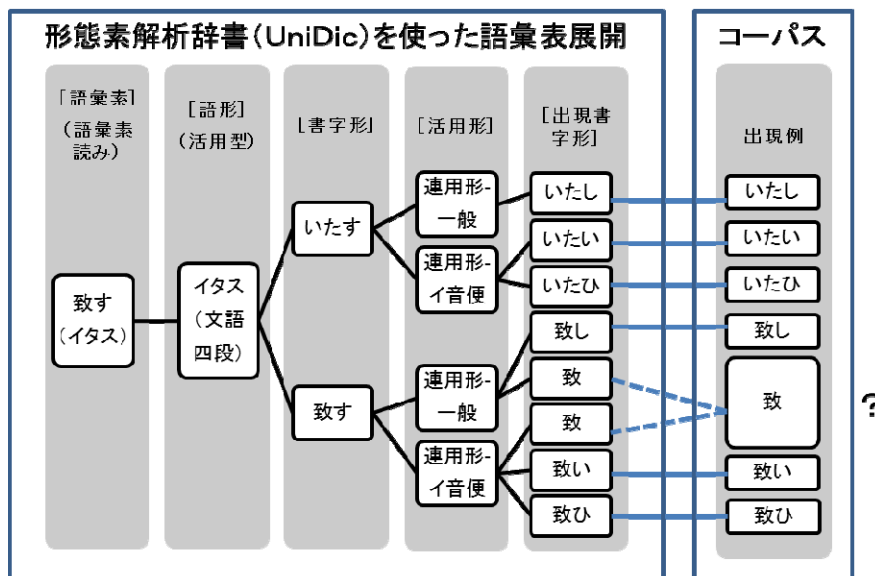


図2 音便形の認定 「致」

『大蔵虎明本狂言集総索引』においても、例えば「い・ふ〔言ふ〕」の項目を見ると、「う (用)」に挙げられた例については、「*印は『云』と漢字表記のため、音便形か

⁴ 蜂谷(1998)も、「狂言台本では、四段活用・ナ行変格活用動詞の連用形が助動詞『た』『たり』、接続助詞『て』などに続く場合」に音便形となることが多い(p.322)とし、サ行四段動詞の音便化については、「そこには語による相違もある程度認められるが、一方、同じような場面で同じ語が原形とイ音便形とで用いられているものもあり、激しい流動の状況をうかがわせる」(p.323)と述べる。

非音便形か不明」(脇狂言)「*印は『云』という漢字表記。そのほとんどはト書きの部分の『云て』の形。『いひて』と読むべきものもあるかもしれない」(聳・山伏類)等の注記があり、やはり活用語尾の表記されていない例については、音便形か非音便形かの判定は難しいことがうかがえる。しかし、コーパスデータにおいては、注をつけることも判断を保留することもできないため、基準を立て、付与する情報を音便形か非音便形かに決定する必要がある。

四段活用動詞のうち、連用形の用例数が多い語(上位10語)について、その語の連用形の全用例数および、活用語尾無表記例の数を整理してみると、表1のような状況であった。

「活用語尾無表記例」は、「云」「参」「申」のように活用語尾が表記されておらず、音便か非音便かの判別が必要な用例の数である。

表1 四段活用動詞連用形の用例数(上位10語)

語	音便の種類	連用形用例数	活用語尾無表記例
言う	ウ音便	1356	814
参る	促音便	563	95
申す	(無)	496	484
持つ	促音便	415	77
取る	促音便	394	63
因る	促音便	358	143
致す	イ音便	343	117
成る	促音便	319	11
思う	ウ音便	298	4
急ぐ	イ音便	288	115

表2 後接語別音便形・非音便形の用例数

語	音便形後接語						非音便形後接語					
	た		たり		て		た		たり		て	
	会話	他	会話	他	会話	他	会話	他	会話	他	会話	他
言う	63	6	31	1	148	99					7	85
参る	139	1	28		161	1	2		16		17	3
申す											1	
持つ	49	1	11		194	22			1		8	14
取る	22	3	9		118	20			3	2	10	31
因る	1				174	11						
致す	55		10		84				4	1	5	
成る	41		5		76	7			7		8	7
思う	23		42		164	2			1		2	
急ぐ	1				146						1	1

また、この10語のうち、活用語尾が表記されている例が「た」「たり」「て」が後接する場合の音便形・非音便形別の用例数を示す(表2)。なお、例えば[出現書字形]を「%

い」「%ひ」「%み」として検索すると、イ音便の仮名表記の例を検索することができ、本調査においてもこのような条件で音便形の仮名表記例を抽出した。

いずれの語においても、全体的には音便形の用例数が非音便形の用例数を上回っており、音便化する例が多いことがうかがえるが、「たり」「て」が後接する場合は、音便形の例・非音便形の例のいずれも見られる。しかし、口語助動詞である「た」が後接する形では、非音便形の例は「まいりた」2例のみで、ほぼ音便形となっていることがわかる。

なお、活用語尾の表記された四段動詞「非音便形+た」の例として、「いだした」（出だす）「思ひ出した」（思い出だす）「かした」（貸す）「おりた」（折る）「作りた」（作る）「たちた」（立つ）等、動詞 18 種、26 例がみられたが、「音便形+た」（動詞 251 種、1654 例）が圧倒的多数である。

このような状況から、四段活用動詞連用形のうち、音便形か非音便形かを確定できない例については、「た」が後接するものについては音便形とし、それ以外の「たり」「て」等が後接するものについては非音便形とした。

ただし、「申す」のように、活用語尾の表記された例がほぼなく、音便形の例が見られない語もある⁵。サ行四段動詞のイ音便形については多くの論考があり、早くは橋本（1962, p.28）に、「中世においても、サ行の動詞の中で、あるものは絶対に音便を起さなかつたことが知られてゐる。召スやオハス或いは申スなどがそれで、中世と言はず古今を通じてこれらの語の音便例は見當らない。」「敬語動詞であることは、音便を起しにくい条件の一つとなる」等の記述があり⁶、この「申」は非音便形と判断すべきものと考えられる。

「申す」とは反対に、「た」が後接する例以外でも非音便形の例があらわれない動詞もある。例えば、「かしこまつて御座る」の「畏まる」、機能語的な「～によつて」「～をもつて」の場合の「困る」「持つ」等である。これらについては定型的な表現として、活用語尾無表記例においても音便形と認定するのが妥当であろう。

このように、音便形もしくは非音便形の例が 1 例もない、あるいは振り仮名が付与されている等の理由から明らかに読みが予想できる例に関しては個別に読みを認定し、判断に迷う語については、「た」が後接するものについては音便形、それ以外の語が後接するものについては非音便形とすることとした。

4. 活用型の認定

『虎明本狂言集』成立期は、活用体系や助動詞語彙の過渡的段階にあたり、それにとまなう問題が、コーパスの語認定においても生じる。一例として、形容詞の活用型の認定の問題を挙げる。

形容詞については、文語活用の終止形である「ーし」と、口語活用の終止形である「ーい」の両形があらわれ、連体形についても同様に、文語活用である「ーき」と、口語活用である「ーい」が見られる。『日本語歴史コーパス』においては、形容詞の[解析活用型]として、「文語形容詞 - ク」「文語形容詞 - シク」および「形容詞」（口語）があり、いずれかの情報を付与する必要がある。文語活用である「ーき」については「文語形容詞」、口語活用である「ーい」については「形容詞」の情報を付与すればよいのだが、例えば「な

⁵ 非音便形の活用語尾が送られている例も「よび【まし】て」の形の 1 例のみである。

⁶ 奥村（1968, pp.44-45）でも、狂言をはじめとする中世末～近世語資料の会話文におけるサ行四段動詞の、全てが音便形の甲型の語および、音便形・非音便形が併存する乙型の語の用例数が整理されている。

がく(長い)」「すずしく(涼しい)」のような、活用語尾が「-く」となっている連用形の例については、文語形容詞、口語形容詞のいずれとするのが妥当であろうか。

表3 形容詞終止形・連体形の活用別用例数

終止形活用語尾	用例数			連体形活用語尾	用例数		
	会話	他	合計		会話	他	合計
口語活用「-い」	725	4	729	口語活用「-い」	1711	16	1727
文語活用「-し」	116	60	176	文語活用「-き」	340	21	361
(活用語尾無表記)	9	3	12	(活用語尾無表記)	12	1	13

表3に、形容詞の終止形と連体形について、活用語尾を口語活用・文語活用にわけ、用例数を示した。終止形・連体形とも、口語活用の語尾の方が優勢であり、『虎明本狂言集』においては形容詞の口語活用化が進んでいたと考えられる。そこで、本コーパスにおける形容詞は、口語活用を原則とし、「-き」「-し」等の明らかな文語活用の例のみ文語活用とすることとした。

ただし、「めでたけれ」「にくけれ」のように、活用語尾が「-けれ」となるものについては、「仮定形」とするか「已然形」とするかが問題となる。室町時代には仮定条件表現は成立しており、『虎明本狂言集』においても、次の例のように、明らかに仮定条件の例があり、必ずしも已然形の已然形たる確定条件の例しか見られないわけではない。

- (4) 又いそぎで【なけれ】ば、某が一細工に致すに依て、来年の今比ならではできませぬよ(仏師 下p.210)

しかし、『虎明本狂言集』には(5)(6)のように「已然形+ど・ども」の形式が残っている。また、(7)のように「こそ」による係り結びも残存しているが、仮に「仮定形」で処理すると、「こそ+已然形」という条件で検索した場合、形容詞がヒットしないことになる。

- (5) 「かほやすがたは【おそろしけれ】ど心はやさしひ(鬼のまま子 下p.489)
 (6) いかによ/\太郎冠者、たらされたは【にくけれ】ども、はやし物がおもしろひ(はりだこ 上p.76)
 (7) 名をとふものこそ【おほけれ】、なぜにみみをとつて引まはずぞ(腹不立 下p.162)

このような点を勘案し、古い形に寄せた「已然形」としておくのが穏当と判断した。

また、本コーパスデータにおいては、動詞は基本的に文語活用としているため、「已然形」とすれば、形容詞に限って「仮定形」があらわれるという例外を避けることができる。さらに、「已然形」としておくことで、既存の「平安時代編」のコーパスと活用形を統一的に検索できる。このように、『虎明本狂言集』および『日本語歴史コーパス』内での統一という点からも、「已然形」とすることとした。

5. おわりに

このように、様々な面で古代語から近代語への過渡的段階である『虎明本狂言集』の言語を現代語および『日本語歴史コーパス』の既存のシステムの中で扱うにあたっては、様々な問題が生じる。蓄積された研究を反映させながら、日本語の史的研究に有用なコーパスを作成することを目指すべきであるが、一方で、語彙や文法事項について、詳細な分類や判別を行ったり、個別の例外を多く認めたりすることにより、ユーザーによる検索や形態素解析辞書の精度維持において、不都合が生じることもある。そのような事情から、便宜的・臨時的な判断を下さざるを得ない面もある。今後の検討や研究の成果によって塗り替えるべき箇所は多く存在するだろうが、これらの問題を解決する手がかりとなりうるのもまた、大量の事例を見渡すことのできるコーパスデータであり、コーパスデータの蓄積が、研究に寄与する部分も大きいと考える。

付 記

本研究は、国立国語研究所共同研究プロジェクト「通時コーパスの設計」(プロジェクトリーダー: 田中牧郎) による成果の一部である。

文 献

- 市村太郎(2014)「近世口語資料のコーパス化—狂言・洒落本のコーパス化の過程と課題—」『日本語学』33-14, pp.96-109
- 大塚光信編(2006)『大蔵虎明能狂言集 翻刻 註解』上・下 清文堂出版
- 奥村三雄(1968)「サ行イ音便の消長」『國語國文』37-1, pp.34-48
- 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』形態論情報規程集第4版(下)」国立国語研究所内部報告書(LR-CCG-10-05-02)
- 小椋秀樹・須永哲矢(2009)『中古和文 UniDic 短単位規程集』科学研究費補助金 基盤研究(C)「和文系資料を対象とした形態素解析辞書の開発」研究成果報告書2(課題番号 21520492)
- 橋本四郎(1962)「サ行四段活用動詞のイ音便に関する一考察」『國語國文』31-4, pp.27-43
- 蜂谷清人(1998)『狂言の国語史的研究—流動の諸相—』明治書院
- 北原保雄・村上昭子(1984)『大蔵虎明本 狂言集総索引 1 脇狂言之類』武蔵野書院
- 北原保雄・鬼山信行(1986)『大蔵虎明本 狂言集総索引 2 大名狂言類』武蔵野書院
- 北原保雄・小川栄一(1982)『大蔵虎明本 狂言集総索引 3 簀類・山伏類』武蔵野書院
- 北原保雄・山崎誠(1989)『大蔵虎明本 狂言集総索引 4 鬼類・小名類』武蔵野書院
- 北原保雄・吉見孝夫(1983)『大蔵虎明本 狂言集総索引 5 女狂言之類』武蔵野書院
- 北原保雄・土屋博映(1984)『大蔵虎明本 狂言集総索引 6 出家座頭類』武蔵野書院
- 北原保雄・大倉浩(1986)『大蔵虎明本 狂言集総索引 7 集狂言之類』武蔵野書院
- 北原保雄・土屋博映(1985)『大蔵虎明本 狂言集総索引 8 万集類』武蔵野書院
- 西尾実・岩淵悦太郎・水谷静夫(編)(2000)『岩波国語辞典 第六版』岩波書店
- 日本国語大辞典 「JapanKnowledge Lib」 <http://japanknowledge.com/library/>
- 室町時代語辞典編修委員会(編)(1985)『時代別国語大辞典 室町時代編一』三省堂

関連 URL

『日本語歴史コーパス』(国立国語研究所) http://www.ninjal.ac.jp/corpus_center/chj/