

# 国立国語研究所学術情報リポジトリ

## 中古和文資料『夜の寢覚』のコーパス構築の試み

メタデータ	言語: Japanese 出版者: 公開日: 2023-11-24 キーワード (Ja): キーワード (En): 作成者: 菊池, そのみ, 菅野, 倫匡 メールアドレス: 所属:
URL	<a href="https://doi.org/10.15084/0002000178">https://doi.org/10.15084/0002000178</a>

## 中古和文資料『夜の寝覚』のコーパス構築の試み

菊池 そのみ (筑波大学)

菅野 倫匡 (筑波大学)

### An Attempt to Build a Corpus of Courtly Literary Works in Early Middle Japanese Language: *Yoru no Nezame*

Kikuchi Sonomi (University of Tsukuba)

Kanno Michimasa (University of Tsukuba)

#### 要旨

本稿は中古和文資料『夜の寝覚』を対象とし、そのコーパスを構築する試みについて作業の手順と現在の進捗状況とを報告した上で今後の展望を述べるものである。まず、本稿では資料として取り上げる『夜の寝覚』の概要を説明し、この資料を対象としてコーパスを構築するための手順を提示した。次に作業を終えた約 2.3 万語のデータに基づき、コーパス全体の総語数を推計すると約 11 万語となることやそれを正解データとすると現行の「中古和文 UniDic」による解析精度が約 98%となることを報告した。また、この正解データに現れた未登録語（未知語）についても併せて一覧として示した。最後に「動詞+テ+動詞」を取り上げ、このようなコーパスを構築することによって容易になる調査の事例を紹介した。

#### 1. はじめに

国立国語研究所による『日本語歴史コーパス』(以下、CHJ)は『同平安時代編』の公開に端を発して段階的に資料や機能の拡充が図られてきた。小木曾(2022:211)によれば、それに伴う一連の構築作業を通して「形態素解析用辞書の整備や本文整備」に関する技術や知見の蓄積が進み、このことによって「研究に必要な本文のコーパスを研究者自身が作ることがすでに可能になってきた」とのことである。

このような昨今の状況を踏まえ、本稿は中古和文資料の1つである『夜の寝覚』を対象とし、そのコーパスを実際に構築する試みについて述べるものである。以下では初めに本稿において取り上げる『夜の寝覚』の概要を示し、次いでコーパスを構築するための手順を説明すると共に現在の進捗状況を報告する。また、このようなコーパスを構築することによって容易になる「動詞+テ+動詞」の調査の事例についても紹介する。

#### 2. 『夜の寝覚』の概要

ここでは『夜の寝覚』の成立、諸本、中古和文資料としての位置づけについて概要を順に示す。なお、資料名は「夜半の寝覚」、「寝覚」、「寝覚物語」などとするものもあるが(阪倉 1964b; 平林 2007)、本稿では『新編日本古典文学全集』に倣って「夜の寝覚」とする。

まず、『夜の寝覚』については11世紀初頭から12世紀初頭までの時期に成立した可能性を指摘する関根・小松(1960)の説や11世紀後半から12世紀前半までの時期に成立したとする阪倉(1964a, 1964b)の説などがあり、概ね中古後期に成立したものと見られる。その作者は『更級日記』の作者として知られる菅原孝標女であると見る説とそれを否定する説とがあり、成立時期と併せて「詳しくは不明」(平林 2007:691)とのことである。なお、成立に関する議論は鈴木(1993)に詳しい。

次に現存する諸本については鈴木(1996)の整理によると五巻本系統に属する写本(島原図書館松平文庫蔵本、天理図書館蔵竹柏園旧蔵本、国立国会図書館蔵本、東北大学狩野文庫蔵本、静嘉堂文庫蔵本、実践女子大学蔵本)と三巻本系統に属する写本(前田家尊経閣文庫蔵本)とに大別し得るとのことである<sup>1</sup>。なお、諸本間の異同は鈴木(1965)の研究や橋本(1933)、高村(1966)、関根・小林・平野・中沢・伊藤・泉(1975)などの校本に譲る。

また、鈴木(1996:583)によれば、諸本のうち「善本と目されるのは島原本と前田家本であり、ともに近世初期の書写である」とのことである。両者のうち前者を底本とするものに『日本古典文学大系』、『日本古典文学全集』、『完訳日本の古典』、『新編日本古典文学全集』があり、後者を底本とするものに関根・小松(1960, 1972)がある。なお、『夜の寝覚』は中間(五巻本における巻二と巻三との間に当たる部分)と末尾(五巻本における巻五の続きに当たる部分)とに欠落があることが知られており、『源氏物語』などに比して「不完全なもの」(永井1960:140)であるとの指摘もある。

続いて中古和文資料における『夜の寝覚』は後期の物語資料の1つに位置づけられ、特に『源氏物語』の影響の著しいもの(平林2007:691)として知られる。また、中古語の語彙や文法の研究においては『日本古典文学大系』に基づいた語彙索引である『夜の寝覚総索引』(阪倉・高村・志水1974)や文字列検索のみを備えた国文学研究資料館『日本古典文学大系本文データベース』を用いることによって研究——山崎(2013)の整理に倣えば、「定性的研究」——が進められてきた(村田2001; 安部・菊池2016; 辻本2016; 菊池2019など)。一方、近年は『日本語歴史コーパス平安時代編』(以下、CHJ平安時代編)の公開・拡充により、それを利用した中古語の研究が増えてきているが、『夜の寝覚』はCHJ平安時代編に収められておらず、コーパスを利用せずには遂行し得ない小林・岡崎(2017)、菊池・菅野(2019)、大川(2020)などの研究——同じく山崎(2013)の整理に倣えば、「定量的研究」——において対象になっていない現状にある。更に前掲の国文学研究資料館『日本古典文学大系本文データベース』が2023年4月1日から公開停止となったことに照らせば、現時点では用例を目視で探すことや索引で引くことを除いて『夜の寝覚』を対象とした研究の実施が困難な状況にあり、定性的研究・定量的研究の双方に資するコーパスの整備が求められるところである。

### 3. コーパス構築の手順

ここではコーパスを構築する際に依拠する資料について述べた上でコーパスを構築するための手順として対象とする資料の本文の電子化と品詞や語種などの情報——小椋(2014)の言うところの「形態論情報」——の認定とを順に説明する。

まず、本稿ではコーパスを構築する際に依拠するものとして『新編日本古典文学全集』を採用することとした。前述の通り、この『新編日本古典文学全集』は鈴木(1996)において「善本」と看做されている近世初期の写本である島原図書館松平文庫蔵本を底本とするものである。なお、コーパスを構築する際に依拠した資料をも仮に「底本」(菅野・菊池2021; 菊池・菅野2022)または「コーパスの底本」(小木曾2022)と呼び得るならば、『新編日本古典文学全集』はコーパスの底本、島原図書館松平文庫蔵本はコーパスの底本の底本ということになる。

<sup>1</sup> 五巻本と三巻本との対応関係については鈴木(1996)も挙げているように前者の巻一・巻二が後者の上巻に当たり、前者の巻三・巻四・巻五が後者の中巻・下巻に当たるとする永井(1960)の指摘がある。

次に対象とする資料の本文を電子化するに当たっては光学文字認識 (OCR) ソフトウェア「読取革命」(Ver.16) を用いて『新編日本古典文学全集』の頭注と現代語訳と (小) 見出しとを除き<sup>2</sup>、本文を電子化した<sup>3</sup>。また、誤りを人手で修正した。

続いて形態論情報の認定に当たっては形態素解析器「MeCab」(Ver.0.996) と解析用辞書「中古和文 UniDic」(Ver.202203) とを用いて形態素解析を実施した。また、各語に認定する形態論情報としては原文文字列、品詞大分類、品詞中分類、品詞小分類、品詞細分類、活用型、活用形、語彙素読み、語彙素表記、書字形出現形、発音形出現形、書字形基本形、発音形基本形、語種、語形、語形基本形、語彙表 ID、語彙素 ID の 18 項目であり、これに参考情報として資料名、巻情報、頁数、行数、位置 ID、ルビの 6 項目を追加し、各語について計 24 項目を認定した<sup>4</sup>。なお、形態論情報については誤りを人手で修正した。

最後にこれまでに述べた作業の手順を模式図として図 1 に示す。

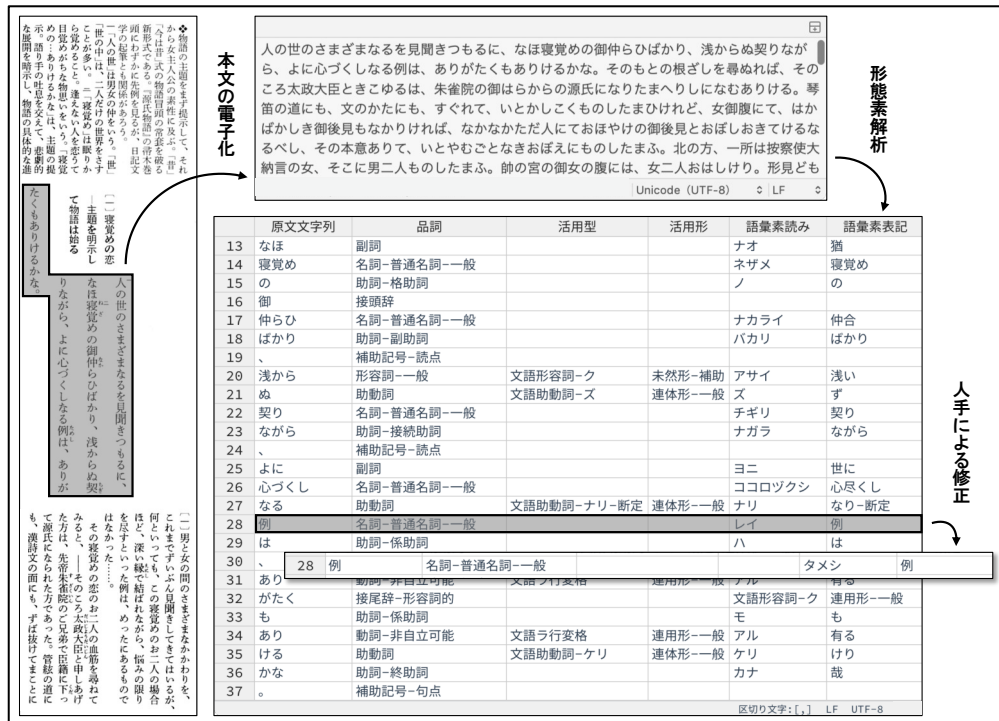


図 1 『夜の寝覚』のコーパスを構築する作業の手順

<sup>2</sup> これ以外にも原則として褐色で示されている文字 (小見出し、注番号、和歌の詠み手、図の説明など) は本文から除いた。

<sup>3</sup> 本稿では『新編日本古典文学全集』の本文を電子化するに当たって公益社団法人著作権情報センターに照会し、当該の行為が「著作物に表現された思想又は感情の享受を目的としない行為」(文化庁著作権課 2019:14) として 2018 年改正の著作権法第 30 条の 4 の権利制限の対象となる——著作権法上の例外に該当する——ことを 2023 年 7 月 26 日に確認した。また、当該のコーパスを公開するに当たっては「軽微性等の要件」(文化庁著作権課 2019:14) ——小木曾 (2023) によると具体的にはキーとなる語の前後 30 語——を満たせば、著作権法第 47 条の 5 第 1 項の準備のための行為として第 47 条の 5 第 2 項における権利制限の対象となることも同日に確認した。

<sup>4</sup> 語形については解析用辞書「中古和文 UniDic」(Ver.202203) では空欄となるが、並行して作業を進めている芥川龍之介賞受賞作品のコーパス——その構想は菅野 (2021) に述べた——と各語に認定する項目を統一して整備するために空欄のままであっても項目としては残してある。

#### 4. 進捗状況

ここでは既に作業を終えた『夜の寝覚』巻一に相当するデータに基づき、コーパス全体の推定規模（推計総語数）、現行の形態素解析の精度、巻一に見られた未登録語（未知語）の一覧を順に示し、進捗状況を報告する。

まず、『夜の寝覚』巻一に相当するデータについて語数——厳密に言えば、短単位数——は 23,185 語であり、これを CHJ 平安時代編所収の各資料の語数と比較すると『大和物語』（26,953 語）に最も近い規模であると言える<sup>5</sup>。また、コーパスの底本である『新編日本古典文学全集』に基づいて巻一のページ数と語数との関係から『夜の寝覚』巻一から巻五までに相当するコーパス全体の総語数を推計すると 113,519 語であり、これは CHJ 平安時代編所収の『大和物語』（26,953 語）と『大鏡』（84,529 語）とを足し合わせた語数に匹敵するものと言える。

次に作業を終えた巻一に相当するデータを正解データとし、現行の解析用辞書「中古和文 UniDic」(Ver.202203) による形態素解析の精度を小木曾（2013）に倣って表 1 に示す。表 1 における「境界」、「品詞」、「語彙素」は小木曾（2013:54-55）と同様に「解析結果において単語の境界が正しかったかどうか」、「境界が正しいことに加えて単語の品詞・活用型・活用形も正しく認定されていたかどうか」、「境界と品詞に加えて語彙素の認定も正しかったかどうか」という各観点からの解析精度を示すものである<sup>6</sup>。

表 1 『夜の寝覚』巻一を正解データとした場合の「中古和文 UniDic」の解析精度

	境界	品詞	語彙素
正解語数	23,185		
出力語数	23,194		
一致語数	23,041	22,856	22,653
再現率	0.99379	0.98581	0.97705
適合率	0.99340	0.98543	0.97668
F 値	0.99360	0.98562	0.97686

表 1 を見ると語彙素認定における解析精度——小木曾（2013:55）によれば、「形態素解析の精度評価では一般に F 値が用いられる」——は F 値が約 98% であり、小木曾（2013）の示す当時のそれが約 97% であることに比して高い水準と言える。また、辞書に未登録の語がデータに含まれる場合に解析精度が低下し、F 値が約 96% になるとする小木曾（2013）の指摘をも踏まえれば、未登録語（未知語）が含まれるデータにおいて高い精度で解析し得たことは特筆に値するものである。これに加えて作品別に見た解析精度について『土佐日記』の解析精度が最も高く語彙素認定で 97.9%、『古今和歌集』が最も低く 93.9% であったとする小木曾（2013:55）の報告に照らしても巻一のデータの解析精度は高い水準であるものと考えられる。

<sup>5</sup> 語数については記号・補助記号・空白を含むことを原則とするが、『夜の寝覚』巻一に相当するデータにおいて記号は 0 短単位であり、補助記号は 3,860 短単位であり、空白は本文の電子化の対象外であることを申し添える。なお、CHJ 平安時代編所収の各資料については言語資源開発センターの Web サイトにおける「語彙統計」(<https://clrd.ninjal.ac.jp/chj/chj-wc.html>) を参照した（2023 年 8 月 10 日確認）。

<sup>6</sup> 表中の「正解語数」、「出力語数」、「一致語数」、「適合率」、「再現率」、「F 値」は小木曾（2013:55）と同様に「人手による修正を行った評価データの語数」、「解析結果として出力されたデータの語数」、「出力語数のうち評価データ（正解）と一致した語数」、「一致語数／出力語数」、「一致語数／正解語数」、「再現率と適合率の調和平均」を指すものである。

このように『夜の寝覚』巻一に相当するデータを高い精度で解析し得たことの背景としては継続的な更新によって解析用辞書「中古和文 UniDic」の精度の向上が図られている点に加えて次の2点が要因となっている可能性がある。まず、小木曾（2022:210）は「コーパスの底本としての本文はさまざまなレベルで異なったものが考えられる」とし、書籍と同様に「読みやすい校訂本文と、原文を尊重した翻字本文」とが役割分担の関係にあるとの前提を示した上でCHJの底本については「多くのユーザーにとって読みやすい新編全集を出発点」としたことを説明した。その上で「形態素解析の障害となる表記上の問題への対処、一般的でない表記が含まれることによる形態素解析精度の低さ」が課題になるとして「コーパスを整備する作業量の面から言えば、原文に近いほど難易度は高い」（小木曾 2022:209）ことを指摘した。この指摘を踏まえると『夜の寝覚』のコーパスの底本である『新編日本古典文学全集』は「校訂を重ねて読みやすくした本文」（小木曾 2022:209）に該当し、「形態素解析の障害となる表記上の問題」（小木曾 2022:209）などが少ないことから解析精度が高い水準になったものと考えられる。次に小木曾（2013:56）は解析用辞書「中古和文 UniDic」による形態素解析の精度について「学習用コーパスの大部分を『源氏物語』が占めるため、全体として『源氏物語』に文体（語彙・語法）に近い作品は高い精度で解析ができる傾向がある」ことを指摘した。この指摘を踏まえれば、本稿において対象としている『夜の寝覚』が中古後期の和文物語資料として『源氏物語』の影響の著しいもの（平林 2007:691）と言われることから解析精度が高い水準になったものと見える。

続いて前掲の解析精度にも大きな影響を与えるものとして『夜の寝覚』巻一に見られた語のうち解析用辞書「中古和文 UniDic」（Ver.202203）に登録のない語——小木曾（2013）によれば、「未登録語（未知語）」——を表2に一覧として示す。

表2 『夜の寝覚』巻一に見られた未登録語の一覧

新語 No.	頁-行 原文文字列	品詞/活用型/活用形	語彙素読み	語彙素表記	書字形基本形	語形基本形	語種	語彙素 ID
1011 027-3	時明	名詞-固有名詞-人名-名	トキアキラ	トキアキラ	時明	トキアキラ	固	
1012 089-14	時明	名詞-固有名詞-人名-名	トキアキラ	トキアキラ	時明	トキアキラ	固	
1021 100-9	わななかしく	形容詞-一般/文語形容詞-シク/連用形-一般	ワナナカシイ	わななかしい	わななかし	ワナナカシ	和	
2011 032-1	おもほれ	動詞-一般/文語下二段-ラ行/連用形-一般	オボホレル	溺ほれる	おもほる	オモホル	和	223745
2021 028-14	しなしなく	形容詞-一般/文語形容詞-シク/連用形-一般	シナジナシイ	品品しい	しなしなく	シナシナン	和	238050
3011 100-5	あち	代名詞	アチラ	彼方	あち	アチ	和	769
3021 077-7	あらはかす	動詞-一般/文語四段-サ行/終止形-一般	アラワカス	表わかす	あらはかす	アラワカス	和	293155
3031 099-1	御殿油	名詞-普通名詞-一般	オオトナブラ	大殿油	御殿油	オオトナブラ	和	257929
3032 103-2	御殿油	名詞-普通名詞-一般	オオトナブラ	大殿油	御殿油	オオトナブラ	和	257929
3041 040-10	垣間見	名詞-普通名詞-一般	カイマミ	垣間見	垣間見	カイバミ	和	195895
3051 058-5	公	名詞-普通名詞-一般	キミ	君	公	キン	和	120142
3061 021-3	年上	名詞-普通名詞-一般	コノカミ	兄	年上	コノカミ	和	233192
3071 042-13	好色	名詞-普通名詞-一般	スキ	数奇	好色	スキ	和	56218
3081 032-12	近勝り	名詞-普通名詞-一般	チカマサリ	近優り	近勝り	チカマサリ	和	284591
3091 019-4	早朝	名詞-普通名詞-副詞可能	ツトメテ	つとめて	早朝	ツトメテ	和	170578
3101 042-2	面杖	名詞-普通名詞-一般	ツラヅエ	頬杖	面杖	ツラヅエ	和	216783
3102 117-12	面杖	名詞-普通名詞-一般	ツラヅエ	頬杖	面杖	ツラヅエ	和	216783
3111 049-1	とほらか	形状詞-一般	トオラカ	遠らか	とほらか	トオラカ	和	309074
3121 079-8	放るかさ	動詞-一般/文語四段-サ行/未然形-一般	ハフラス	放らかす	放るかす	ハフラス	和	248531
3131 051-6	ひとびとしく	形容詞-一般/文語形容詞-シク/連用形-一般	ヒトビトシイ	人人しい	ひとびとし	ヒトビトシ	和	249452
3141 050-2	行明	名詞-固有名詞-人名-名	ユキアキラ	ユキアキラ	行明	ユキアキラ	固	340807
3151 027-2	行頼	名詞-固有名詞-人名-名	ユキヨリ	ユキヨリ	行頼	ユキヨリ	固	345242
3152 028-6	行頼	名詞-固有名詞-人名-名	ユキヨリ	ユキヨリ	行頼	ユキヨリ	固	345242
3153 029-12	行頼	名詞-固有名詞-人名-名	ユキヨリ	ユキヨリ	行頼	ユキヨリ	固	345242
3154 030-4	行頼	名詞-固有名詞-人名-名	ユキヨリ	ユキヨリ	行頼	ユキヨリ	固	345242
3161 096-11	緩し	動詞-一般/文語四段-サ行/連用形-一般	ユルス	許す	緩す	ユルス	和	38938
3171 025-2	よかん	形容詞-非自立可能/文語形容詞-ク/連体形-撥音便	ヨイ	良い	よし	ヨシ	和	36983
3181 035-2	由々し	形容詞-一般/文語形容詞-シク/終止形-一般	ヨシヨシイ	由由しい	よしよし	ヨシヨシシ	和	247943

なお、形態素解析用辞書「UniDic」は語彙素>語形>書字形>発音形の4段階の階層構造を持つことから必要な語を各階層に追加し得るとする小木曾(2013)の指摘を踏まえ、この階層に即して整理すると『夜の寝覚』巻一における未登録語は語彙素の階層に登録すべき語が異なり2語(延べ3語)、語形の階層に登録すべき語が異なり2語(延べ2語)、書字形の階層に登録すべき語が異なり19語(延べ23語)であり、最終的に合計すると異なり23語(延べ28語)となる。

以下では表2に示した『夜の寝覚』巻一における未登録語のうち「わななかしく」(新語 No.1021)、「おもほれ」(新語 No.2011)、「しなしなしく」(新語 No.2021)、「よかん」(新語 No.3171)の4語を取り上げ、各語について説明を加えることとする<sup>7</sup>。

まず、(1)の下線部は解析用辞書「中古和文 UniDic」に登録のない語と見られる。この語について『日本国語大辞典』第2版は「[形シク] (動詞「わななく(戦慄)」の形容詞化) 恐れや緊張などのため、ふるえが起りそうになるほどである。」とし、(1)の用例のみを掲げている。このことから察するに現行の解析用辞書「中古和文 UniDic」に該当する語の登録がないのは CHJ 平安時代編所収の和文資料に用例が見られないからであるという可能性が考えられる。これを踏まえて《わななかしい》を語彙素の階層に登録すべき語であると判断した。

- (1) …御帳の外にみざり出でて、ものなど言ひ紛らはし、さらぬ顔にと思ふ心地も、いとわななかしくわびしけれど、とかく聞き言ふべきやうもなければ…  
(巻一、100頁9行目<sup>8</sup>)

次に(2a)の下線部は解析用辞書「中古和文 UniDic」に登録のない語と見られる。この語について『日本国語大辞典』第2版は「[自ラ下二] おぼれる。涙にくれる。おぼほる。」として(2a)の用例を最初に掲げており、『新編日本古典文学全集』の頭注は「「おもほれ」は「おぼほれ」の誤りか。」と説明している。このことから考えると(2a)の下線部は既に解析用辞書「中古和文 UniDic」に登録のある《おぼれる》——実際に CHJ 平安時代編においては(2b)のように〈オボホル〉として現れる——と同じ語彙素の異なる語形である可能性がある。また、CHJ 平安時代編において(2c)の下線部と(2d)の下線部とが共に同じ語彙素《守る》であることも踏まえ、〈オモホル〉を語彙素《おぼれる》の語形の階層に登録すべき語であると判断した。

- (2) a. …と思ひつづくるに、あたらしう、口惜しく、涙におもほれまどひながらも、思ひやりいと静かなる人にて、…  
(巻一、32頁1行目)  
b. …いとといみじきに、聞こえん方なきことどもなれば、ただ涙におぼほれたるばかりをかごとにて、はかばかしうも答へやらずなりぬ。  
(源氏物語・蜻蛉、215頁、20-源氏 1010\_00052・52270)

<sup>7</sup> 説明に当たっては特に書字形を隅つき括弧で括り、語形を山括弧で括り、語彙素を二重山括弧で括って便宜的に区別した。

<sup>8</sup> 用例の引用に当たって縦書きを横書きに改め、ルビ等を除いた。また、下線は強調の意を示し、「…」は文中の一部を省略した意を示すために引用者が施したものである。なお、出典は用例の末尾に示し、『夜の寝覚』巻一の場合には『新編日本古典文学全集』の頁番号・行番号を併記し、CHJ 所収の資料の場合には資料名と『新編日本古典文学全集』の頁番号と CHJ における位置 ID・開始位置とを併記した。

- c. …かしこにまぼりでものせむ、世の中いとはかなければ、いまはかたちをもこと  
 になしてむとてなむ… (蜻蛉日記・下, 282 頁, 20-蜻蛉 0974\_00009・44200)
- d. うち見おこせて、つくづくうちまもりて、いといみじと思ひたり。とまるはさらにも  
 いはず。 (蜻蛉日記・上, 141 頁, 20-蜻蛉 0974\_00002・71710)

続いて (3a) の下線部は解析用辞書「中古和文 UniDic」に登録のない語と見られる。この語について『日本国語大辞典』第2版は立項しておらず、『新編日本古典文学全集』の頭注は「「しなしなし」は、いかにも上品である意。」と説明している。また、これと清濁の対立を持つように見える「しなじなし」について『日本国語大辞典』第2版は「[形シク] 素性がいい。上品である。けだかい。」と説明している。このことから察するに (3a) の下線部は既に解析用辞書「中古和文 UniDic」に登録のある《品品しい》——実際に CHJ 平安時代編においては (3b) のように〈シナジナシ〉として現れる——と同じ語彙素の異なる語形である可能性が考えられる。これを踏まえれば、〈シナシナシ〉を語彙素《品品しい》の語形の階層に登録すべき語であると判断した。

ただし、清濁の対立を持つ語形の登録については (3c) に掲げる規定があり、この規定に照らすと『日本国語大辞典』第2版における「しなじなし」の項に「古くは清音であったとの記述」のないことから〈シナシナシ〉を《品品しい》の語形の階層に登録し得ないものと解せる余地もあるという点には留意する必要がある。

- (3) a. 頭つき、様体ほそやかに、しなしなしく、きよらなるに、髪のいとつややかに  
 ゆるゆるとかかりて、目やすき人かな、と見ゆるに、… (巻一, 28 頁 14 行目)
- b. さては、扇よりかみの額つきぞ、あやしく人のかたちを、しなじなしくも下りて  
 も、もてなすところなんめる。(紫式部日記, 156 頁, 20-紫式 1010\_00001・115850)
- c. 仮名表記の出現形に対し、現代語で濁音化するところに濁点がついていないという  
 場合の処理に関しては、『日本国語大辞典』第2版で古くは清音であったとの記述  
 が確認できる限り、別語形として清音形を登録し、清音で読む。

(国立国語研究所コーパス開発センター (池上尚) 2016:141, 3.1.6.清濁)

最後に (4a) の下線部は解析用辞書「中古和文 UniDic」に登録のない語と見られる。この語を含む「よかんめり」について『新編日本古典文学全集』の現代語訳は「よいようだ」としており、「よかん」に助動詞「めり」が下接したものと解し得る。このことから考えると (4a) の下線部は既に解析用辞書「中古和文 UniDic」に登録のある《良い》の連体形-撥音便——実際に CHJ 平安時代編においては (4b) のように語形〈ヨカン〉が書字形【よか】として現れる——と同じ語彙素の同一の語形の異なる書字形である可能性がある。また、CHJ 鎌倉時代編においては (4c) のように《良い》の語形〈ヨカン〉が書字形【よかん】として現れており、CHJ 平安時代編においては (4d) のように《繁い》の語形〈シゲカン〉が書字形【しげかん】として現れている。これを踏まえて【よかん】を《良い》の語形〈ヨカン〉の書字形の階層に——厳密には書字形基本形【よし】の書字形出現形【よかん】として——登録すべき語であると判断した。

- (4) a. …ただこの御身に添ひて起き臥したまふを、大臣も、「いとよかんめり。この御方  
 にあづかりて、おぼしうしろめ」とて、… (巻一, 25 頁 2 行目)



- b. …いとをかしげにひねり縫ひたまひければ、「いとよかめり。ことなるかほかたちなき人は、物まめやかに習ひたるぞよき」とて…  
 (落窪物語・巻之一, 19 頁, 20-落窪 0986\_00001・7780)
- c. …「山里は物のさびしき事こそあるなれども、世のうきよりは住みよかんなるものを」とて、おぼしめしたたせ給ひけり。  
 (平家物語・平家灌頂巻・大原入, 506 頁, 30-平家 1250\_13002・2820)
- d. またの日も、まだしきに、「昨日は、うそぶかせたまふこと、しげかんめりしかば、えものも聞こえずなりにき。(蜻蛉日記・下, 345 頁, 20-蜻蛉 0974\_00012・10450)

## 5. 調査の事例

ここでは『夜の寝覚』のコーパスを構築することによって容易になる調査の事例について紹介する。具体的には「動詞+テ+動詞」——大木 (2022) に倣えば、「V1+テ+V2<sup>9</sup>」——について CHJ を対象とした大木 (2022) の調査結果と『夜の寝覚』巻一を対象とした調査結果とを比較する。

なお、調査に当たってはプログラミング言語「Python」(Ver.2.7.18) を用いてコーパスを検索するアプリケーションを作成し、用例を抽出した。このコーパス検索アプリケーションは図 2 のように検索条件を指定すると図 3 のように検索結果を表示する機能を備えたものである。

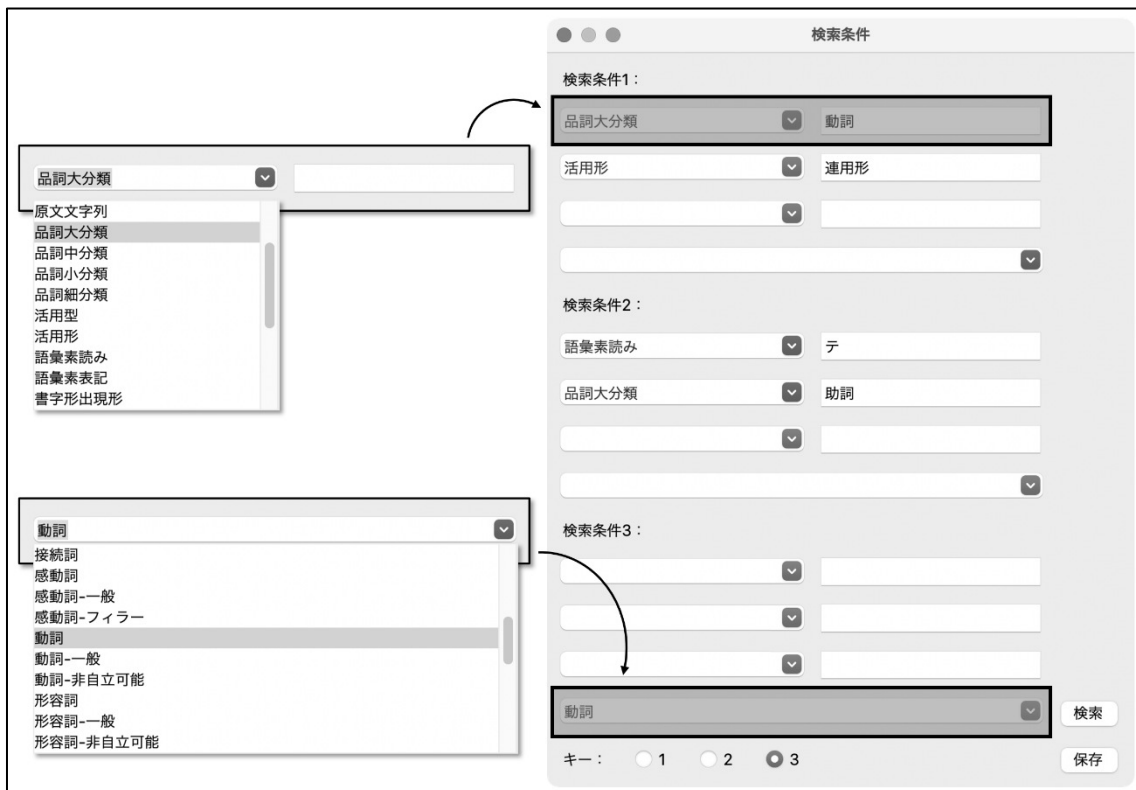


図 2 検索アプリケーションにおける検索条件の設定画面

<sup>9</sup> 大木 (2022:66) は「テ形補助動詞」の成立について述べた青木 (2020) を引用しつつ「動詞テ形に動詞が後接する場合の前項動詞を V1、後項動詞を V2 とする」と説明している。

検索結果

検索結果：108件

巻情報	頁数	行数	位置ID	前文脈	原文文字列	後文脈	品詞	活用型	活用形	語彙素読み	語彙素表記
巻1	17	6	6610	君 の い と すぐ れ て	生ひ	たち たまふ に は 、	動詞-一般	文語上二段-八行	連用形-一般	オイル	生いる
巻1	17	12	7630	の 年 を 経 て	弾き	しみ たる より も 、	動詞-一般	文語四段-カ行	連用形-一般	ヒク	弾く
巻1	17	14	8060	に や が て 傾 き か かり て	御殿籠り	入り たる に 、 小	動詞-一般	文語四段-ラ行	連用形-一般	オオントノ	御殿籠る
巻1	18	6	9610	、 琵琶 を 持 て	来	て 、 「今宵 の	動詞-非自立可能	文語力行変格	連用形-一般	クル	来る
巻1	18	10	10290	の 今宵 下 り 来 て	教へ	た て ま つ ら む 」 と て 失 せ	動詞-一般	文語下二段-八行	連用形-一般	オシエル	教える
巻1	19	2	11310	琵琶 を 取 り 寄 せ て	弾き	たまふ に 、 大臣 聞 き	動詞-一般	文語四段-カ行	連用形-一般	ヒク	弾く
巻1	19	10	12890	し 月 日 を 数 へ て	待つ	に 、 また の 年	動詞-一般	文語四段-タ行	連体形-一般	マツ	待つ
巻1	19	11	13160	の ある か ぎ り 出 だ し て	弾き	たまへ れ ば 、 殿	動詞-一般	文語四段-カ行	連用形-一般	ヒク	弾く
巻1	20	15	16140	、 これ こ そ すぐ れ て	聞こゆれ	。 昔 より と り わ き 殿	動詞-非自立可能	文語下二段-ヤ行	已然形-一般	キコエル	聞こえる
巻1	24	3	25880	の 風 に 合 は せ て	弾き	たまへ る 音 の 、	動詞-一般	文語四段-カ行	連用形-一般	ヒク	弾く
巻1	25	1	28220	の 君 と き こ え て	ものし	たまふ 。 こ の 殿	動詞-非自立可能	文語サ行変格	連用形-一般	モノスル	ものする
巻1	25	11	29810	の 御身 に 添 ひ て	起き	臥 した まふ を 、 大臣	動詞-一般	文語上二段-カ行	連用形-一般	オキル	起きる
巻1	27	3	34090	は 、 所 さ り て	忌み	たまふ べき な れ ば 、	動詞-一般	文語四段-マ行	連用形-一般	イム	忌む
巻1	27	8	35000	女 な む 、 渡 り て	住み	さ ぶ ら ふ な り 。 月 明 き	動詞-一般	文語四段-マ行	連用形-一般	スム	住む
巻1	27	13	36160	弁 少 將 に 契 り て	かしづき	さ ぶ ら ふ 三 に あ た る は	動詞-一般	文語四段-カ行	連用形-一般	カシズク	傅く
巻1	28	1	36680	弁 少 將 に 契 り て	さぶらふ	な り 」 と 申 せ ば	動詞-一般	文語四段-八行	連体形-一般	サブラウ	侍う
巻1	28	3	37060	に 歩 み 寄 り た ま ひ て	聞き	たまへ ば 、 琴 の	動詞-一般	文語四段-カ行	連用形-一般	キク	聞く
巻1	28	11	38530	は 、 い と すぐ れ て	聞こゆ	。 こ な た も か な た も	動詞-非自立可能	文語下二段-ヤ行	終止形-一般	キコエル	聞こえる
巻1	30	5	43030	も と に 伝 ひ 寄 り て	見	たまへ ば 、 池 、	動詞-非自立可能	文語上一段-マ行	連用形-一般	ミル	見る
巻1	31	1	44610	て 、 閉 り 入 り て	見	たまへ ば 、 和 琴 の	動詞-非自立可能	文語上一段-マ行	連用形-一般	ミル	見る
巻1	32	4	47760	。 人 氣 に お ど る き て	見	返 り た る ほ ど に 、	動詞-非自立可能	文語上一段-マ行	連用形-一般	ミル	見る
巻1	36	13	59620	際 は な ほ 忍 び て	やみ	な む 」 と 思 ふ	動詞-一般	文語四段-マ行	連用形-一般	ヤム	止む
巻1	37	4	60690	明 け 暮 れ 出 だ し 入 れ て	見	つ べ か ら む 人 に	動詞-非自立可能	文語上一段-マ行	連用形-一般	ミル	見る
巻1	37	6	61240	、 中 宮 に 申 し て	召し	取 ら せ た て ま つ り て 、	動詞-一般	文語四段-サ行	連用形-一般	メス	召す
巻1	37	10	62020	さ る 私 物 に 忍 び て	見	む 。 さ て は 人	動詞-非自立可能	文語上一段-マ行	連用形-一般	ミル	見る
巻1	37	13	62450	の う ち に 忍 び て	通は	む こ そ 、 い と 見 苦 し か る	動詞-一般	文語四段-八行	未然形-一般	カヨウ	通う
巻1	38	14	65300	の う ち へ 入 り て	見	たまへ ば 、 げ に 物 忌	動詞-非自立可能	文語上一段-マ行	連用形-一般	ミル	見る
巻1	39	3	65910	」 と 、 せ め て	思ひ	く た し 思 ひ 覺 ます に も	動詞-一般	文語四段-八行	連用形-一般	オモウ	思う
巻1	39	12	67600	、 筆 に ま か せ て	書き	流 さ れ た る は 、	動詞-一般	文語四段-カ行	連用形-一般	カク	書く

図3 検索アプリケーションにおける検索結果の表示画面

まず、前述した大木（2022）の調査について概要を述べる。大木（2022:67）は「テ形補助動詞」の成立について検討するための出発点として「V1+テ+V2」に「どのようなものがあるのか、また、それが概略どのように移り変わってきているのか」という点を確認するとしてCHJ所収の資料を対象に「平安時代」から「江戸時代」までの「V1+テ+V2」を抽出した上でV2に着目し、その出現頻度の上位30位までの語を一覧として示している。

これに倣って本稿においても既にコーパス構築の作業を終えている『夜の寝覚』巻一から「V1+テ+V2」の用例を抽出した。具体的には前述した検索アプリケーションにおいて図2のように検索条件1の品詞大分類を「動詞」とした上で活用形を「連用形」とし、検索条件2の語彙素読みを「テ」とした上で品詞大分類を「助詞」とし、検索条件3の品詞大分類を「動詞」とし、検索条件3をキーとして検索し、用例を抽出した<sup>10</sup>。その上でV2を出現頻度の順に並べて大木（2022:68-69）の提示する表1における「平安時代」の結果と共に表3に示す<sup>11</sup>。なお、V2を挙げる際には大木（2022）の結果と比較するための便宜を図り、大木（2022）に倣って語彙素（現代語の動詞の形）を示すこととする<sup>12</sup>。

<sup>10</sup> 図2に示すように検索条件として品詞を指定する場合は品詞大分類などを個別に指定する方法——図2の検索条件1——と一括して指定する方法——図2の検索条件3——との2通りがあるが、得られる結果は同じものである。

<sup>11</sup> 大木（2022）の表1における網掛けと丸印とを除いて示した。また、大木（2022:87）には『日本語歴史コーパス』バージョン2021.3』を利用したことや当時のCHJ平安時代編所収の資料を対象としたことなどについて記載はあるが、調査の対象とした資料の詳細や用例を抽出する際の検索条件については具体的な言及がないことを申し添える。

<sup>12</sup> 大木（2022:67）も述べているように語彙素表記の「居る」には「ゐる」「をり（る）」が含まれていることから語彙素表記では語彙素読みの異なる語を弁別し得ないという点には留意する必要がある。

表3 「V1+テ+V2」におけるV2

大木(2022)の調査結果			本稿における『夜の寝覚』巻一のデータ					
順位	語彙素表記	用例数	順位	語彙素表記	用例数	順位	語彙素表記	用例数
1	見る	426	1	見る	16	16	引く-他動詞	1
2	侍る	257	2	思う	9	16	起こす	1
3	参る	222	3	居る	6	16	隠す	1
4	来る	206	4	聞く	4	16	わななく	1
5	出でる	166	4	弾く	4	16	止む	1
6	居る	152	6	侍る	3	16	待つ	1
7	行く	145	6	思す	3	16	窺う	1
8	おわす	139	6	起きる	3	16	応える	1
9	有る	138	6	立つ	3	16	沿う	1
10	思う	114	10	問う	2	16	眺める	1
11	聞こえる	113	10	聞こえる	2	16	住む	1
12	言う	110	10	忍びる	2	16	独り言つ	1
13	奉る	109	10	有る	2	16	出だす	1
14	読む	107	10	侍う	2	16	念ずる	1
15	立つ	101	10	出でる	2	16	籠もる	1
16	返る	88	16	背く	1	16	伏す	1
17	宣う	83	16	淡める	1	16	おわす	1
18	往ぬ	80	16	来る	1	16	ものする	1
19	渡る	80	16	見える	1	16	召す	1
20	思す	74	16	通う	1	16	訪れる	1
21	おわします	73	16	為る	1	16	上る	1
22	ものする	73	16	返る	1	16	給う-尊敬	1
23	見える	72	16	御殿籠る	1	16	寝る	1
23	聞く	72	16	傳く	1	16	頼む	1
25	伏す	70	16	抗う	1	16	生いる	1
26	遣る	67	16	知る	1	16	惑う	1
27	入る	65	16	行く	1	16	然る	1
28	侍う	63	16	覚える	1	16	教える	1
29	泣く	60	16	仕る	1	16	耐える	1
30	止む	56	16	忌む	1	16	書く	1
30	書く	56					計	108

次に表3に基づいて大木(2022)におけるCHJ平安時代編所収の資料を対象とした調査結果と本稿における『夜の寝覚』巻一を対象とした調査結果との比較を試みる。両者の関係を捉えるために表3に示した動詞の出現順位についてSpearmanの順位相関係数を利用して相関分析を実施した。その結果、共通して見られる動詞(20語<sup>13</sup>)について大木(2022)の順位と巻一の順位との間に相関は認められなかった( $\rho = .386, p = .093, n.s.$ )。このように両者に相関が認められなかったことは『夜の寝覚』が『源氏物語』の影響の著しいもの(平林2007:691)と言われてきたことを踏まえれば、CHJ平安時代編所収の和文資料と比較して『夜の寝覚』が特異であることを示唆するものとは看做せず、調査の規模が小さいことの影響を受けている可能性が考えられる。

また、表3について両者の共通点を具体的に見ると大木(2022)は上位30位に見られる動詞のうち「見る」、「来る」、「居る」、「行く」、「有る」、「遣る」が現代語におけるテ形補助動詞のV2として見られることを示しており、これらの動詞は『夜の寝覚』巻一にも用例が見られた。特に「見る」は大木(2022)の調査結果においても『夜の寝覚』巻一の調査結果においても最も出現頻度の高い動詞であったという点は注目し得る。更に大木(2022)は現代語と異なる点として中古和文資料においては「侍る」、「おわす」、「有る」などの「敬語

<sup>13</sup> この20語については表3において網掛けにして示したものである。

系動詞」が V2 として見られることを指摘しており、これらが『夜の寝覚』巻一においても同様に見られるものであることも確認し得た。

一方、両者の相違点についても具体的に見てみると大木（2022）の示す上位 30 位の動詞には見られないものの『夜の寝覚』巻一において上位——具体的には 4 位——に現れる動詞として「弾く」が挙げられる。これは『夜の寝覚』巻一において琵琶や琴を弾く様子が繰り返し描かれることに起因するものと見える<sup>14</sup>。また、『夜の寝覚』巻一においては見られないものの大木（2022）の示す上位 30 語の動詞——具体的には 3 位——に現れるものとしては「参る」が挙げられる。この「参る」が『夜の寝覚』巻一において V2 として現れない理由は定かでないが、物語の展開や他の動詞との関係について更に検討を深める余地のあるものと見える<sup>15</sup>。当然、「V1+テ+V2」の文法的な振る舞いを明らかにするためには複数の資料から多くの用例を収集（し、資料の個別性を捨象することによって得られる資料間の共通性を記述）することが求められるが、個別の資料に着目して当該の資料の特徴を明らかにすることも同様に重要であり、そのためには両者の相違点を見ることも必要になると言える。

このように本稿において構築を進めている『夜の寝覚』のコーパスは CHJ 平安時代編と同様に形態論情報を認定していることから索引の利用や文字列検索に比して品詞や活用形を指定した検索を容易にするものであると同時に CHJ 平安時代編を対象とした調査結果と比較することをも可能にするものであると言える。

## 6. おわりに

本稿では中古和文資料『夜の寝覚』のコーパスを構築する試みについて述べてきた。この『夜の寝覚』のコーパスが完成すれば、中古語を対象とした語彙や文法に関する定性的研究と共に宮島（1970）や小林・小木曾（2013）などに類する定量的研究も可能になり、未だに定説を見ない『夜の寝覚』の成立に関する定量的な分析を実施するための環境も整うものと考えられる<sup>16</sup>。これに関連して『夜の寝覚』の異なる底本間の比較や『夜の寝覚』と改作本と言われる『夜寝覚物語』との比較を実施することも展望として挙げられる<sup>17</sup>。

<sup>14</sup> そもそも『夜の寝覚』巻一において「弾く」は 22 例であり、巻一における動詞の合計 4,472 例のうち 0.49% を占める動詞である。一方、CHJ 平安時代編所収の和文資料において「弾く」は 133 例であり、動詞の合計 187,962 例のうち 0.07% を占める動詞である。このことから『夜の寝覚』巻一では他の中古和文資料に比して「弾く」が多く用いられている可能性が考えられる。

<sup>15</sup> ただし、試みに『夜の寝覚』巻二を見てみると V2 として現れる「参る」が数例は認められることから『夜の寝覚』に全く用例がないとは言えず、巻一に用例がないのは偶発的なものである可能性もある。

<sup>16</sup> 既に『夜の寝覚』の作者については定量的な分析を試みた北原（2015）の研究があるが、『日本古典文学大系』の本文を用いているという点において本稿の構築しているコーパスとは底本が異なることに加えて『夜の寝覚』、『浜松中納言物語』、『更級日記』、『紫式部日記』を比較することに留まっており、本来は他の資料との比較によって相対的に捉えられるはずの資料間の類似性について更なる検討の余地を残すものと考えられる。これに関連して菊池・菅野（2019）では品詞の構成比率の観点から中古和文資料の類似性について韻文と散文との比較や資料間の比較を実施しており、これに『夜の寝覚』を含めて同様に検討することによって明らかになる中古和文資料における位置づけを踏まえて『夜の寝覚』の成立に関する議論を進められる可能性がある。また、いずれも菅原孝標女によって書かれたとも言われる『夜の寝覚』、『浜松中納言物語』、『更級日記』の作者説について検討するに当たっては『浜松中納言物語』のコーパスを構築することも併せて必要になる。

<sup>17</sup> 小木曾（2022:210）も「伝本が複数ある場合には、複数の本文をコーパスとして比較、対照できるように整備することが求められる」とし、同一の資料について底本の異なるコーパスの構築を提案している。

なお、昨今では「研究者各人が自らの研究課題に即したコーパスを用意し分析に利用することも容易になりつつある」（間淵 2020:121）と言われるが、コーパスを自らの手によって用意（構築）することは困難を伴うものと言わざるを得ず、実際にはその「開発を牽引する国立国語研究所」（間淵 2020:114）の手を借りなければ、依然として実現し難い現状にあるものと見える。本稿のような試みの蓄積によって「国語研の外部の研究者であっても歴史的資料のコーパスを作ることができる環境」（小木曾 2022:211）の整備が更に進展することが望まれるところである。

### 謝 辞

本研究は JSPS 科研費 JP22K19985, JP22K19986 の成果の一部である。なお、作業の実施に当たっては黒田優月氏（筑波大学学生）の助力を得た。

### 参 考 文 献

- 青木博史（2020）「動詞連用形＋動詞」から「動詞連用形＋テ＋動詞」へ―「補助動詞」の歴史・再考― 青木博史・小柳智一・吉田永弘（編）『日本語文法史研究 5』, pp.197-226, ひつじ書房.
- 安部清哉・菊池そのみ（2016）「男が泣く日本文学の系譜を探って―男主人公が最も泣く『夜の寝覚』―」『学習院大学国語国文学會誌』 59, pp.76(57)-59(74), 学習院大学文学部国語国文学會.
- 大木一夫（2022）「テ形補助動詞成立史概略、拾遺」青木博史・岡崎友子・小木曾智信（編）『コーパスによる日本語史研究 中古・中世編』, pp.65-87, ひつじ書房.
- 大川孔明（2020）「叙述語から見た平安鎌倉時代の文学作品の文体類型」『計量国語学』 32:6, pp.331-345, 計量国語学会.
- 小木曾智信（2013）「中古仮名文学作品の形態素解析」『日本語の研究』 9:4, pp.49-62, 日本語学会.
- 小木曾智信（2022）『日本語歴史コーパス』 中古・中世のデータ構築について」青木博史・岡崎友子・小木曾智信（編）『コーパスによる日本語史研究 中古・中世編』, pp.199-214, ひつじ書房.
- 小木曾智信（2023）『昭和・平成書き言葉コーパス』の構築・公開と権利処理」『日本語学会 2023 年度春季大会予稿集』, pp.145-147, 日本語学会.
- 小椋秀樹（2014）「形態論情報」前川喜久雄（監）山崎誠（編）『書き言葉コーパス―設計と構築―』, pp.68-88, 朝倉書店.
- 菅野倫匡（2021）「芥川賞作品コーパスの構築に向けて―語彙調査に関する未解決の問題との関連から―」『日本語と日本文学』 67, pp.75-88, 筑波大学日本語日文学会.
- 菅野倫匡・菊池そのみ（2021）「和歌集における品詞の構成比率の算出方法に関する試論―計量語彙論のための標本抽出法を中心に―」『計量国語学』 33:3, pp.162-177, 計量国語学会.
- 菊池そのみ（2019）「古代語の「ての」について」『筑波日本語研究』 23, pp.113-134, 筑波大学人文社会科学研究所日本語学研究室.
- 菊池そのみ・菅野倫匡（2019）「勅撰和歌集の語彙の量的構造をめぐって―品詞の構成比率の観点から―」『国語語彙史の研究』 38, pp.360(57)-336(81), 和泉書院.

- 菊池そのみ・菅野倫匡（2022）「和歌語彙の研究に『日本語歴史コーパス』を利用する際の留意点」『筑波日本語研究』26, pp.37-62, 筑波大学人文社会科学研究所日本語学研究室.
- 北原慈子（2015）「古典文学研究における計量文献学的手法をめぐって—『更級日記』『浜松中納言物語』『夜半の寢覚』『紫式部日記』を題材として—」『Studies in Language Science』5, pp.75-103, 立命館大学大学院言語教育情報研究科.
- 国立国語研究所コーパス開発センター（池上尚）（編）（2016）『『日本語歴史コーパス平安時代編』形態論情報規程集』, 国立国語研究所.
- 小林雄一郎・岡崎友子（2017）「中古における接続表現の統計的分析—指示詞を構成要素とするものを中心に—」『国立国語研究所研究論集』13, pp.65-77, 国立国語研究所.
- 小林雄一郎・小木曾智信（2013）「中古和文における個人文体とジャンル文体—多変量解析による歴史的資料の文体研究—」『国立国語研究所論集』6, pp.29-43, 国立国語研究所.
- 阪倉篤義（1964a）「『夜の寢覚』の文章」『國語と國文學』41:10, pp.144-156, 東京大学国語国文学会.
- 阪倉篤義（1964b）「解説」阪倉篤義（校注）『日本古典文学大系 78 夜の寢覚』, pp.3-37, 岩波書店.
- 阪倉篤義・高村元継・志水富夫（編）（1974）『夜の寢覚総索引』, 明治書院.
- 鈴木一雄（1993）「『夜の寢覚』と『更級日記』の作者」鈴木一雄『王朝女流日記論考』, pp.258-311, 至文堂.
- 鈴木一雄（1996）「解説」鈴木一雄（校注）『新編日本古典文学全集 28 夜の寢覚』, pp.561-589, 小学館.
- 鈴木弘道（1965）『寢覚物語の基礎的研究』, 塙書房.
- 関根慶子・小松登美（1960）『寢覚物語全釈』, 學燈社.
- 関根慶子・小松登美（1972）『増訂寢覚物語全釋』, 學燈社.
- 関根慶子・小林太枝子・平野由紀子・中沢礼伊子・伊藤弘子・泉民子（1975）「寢覚物語對校」関根慶子教授退官記念会（編）『関根慶子教授退官記念寢覚物語對校・平安文学論集』, pp.1-485, 風間書房.
- 高村元継（1966）『校本夜の寢覚』, 明治書院.
- 辻本桜介（2016）「主節主体の動きを表す動詞終止形に接続するトテについて—引用と異なる機能の分析—」『日本語の研究』12:2, pp.35-51, 日本語学会.
- 永井和子（1960）「『ねざめ』の構造」『平安文学研究』25, 平安文学研究会.
- 橋本佳（1933）『校本夜半の寢覚』, 大岡山書店.
- 平林文雄（2007）「夜の寢覚」飛田良文・遠藤好英・加藤正信・佐藤武義・蜂谷清人・前田富祺（編）『日本語学研究事典』, p.691, 明治書院.
- 文化庁著作権課（2019）『デジタル化・ネットワーク化の進展に対応した柔軟な権利制限規定に関する基本的な考え方（著作権法第30条の4, 第47条の4及び第47条の5関係）』, [https://www.bunka.go.jp/seisaku/chosakuken/hokaisei/h30\\_hokaisei/pdf/r1406693\\_17.pdf](https://www.bunka.go.jp/seisaku/chosakuken/hokaisei/h30_hokaisei/pdf/r1406693_17.pdf)（2023年7月26日確認）, 文化庁.
- 間淵洋子（2020）「数理的研究」『日本語の研究』16:2, pp.114-121, 日本語学会.
- 宮島達夫（1970）「語いの類似度」『国語学』82, pp.42-64, 国語学会.
- 村田菜穂子（2001）「平安時代の形容動詞—〜ゲナリと〜カナリ—」『国語学』52:1, pp.16-30, 国語学会.

山崎誠 (2013) 「形式語研究の方法論—定性的研究と定量的研究—」 藤田保幸 (編) 『形式語研究論集』, pp.1-18, 和泉書院.

### 資料

- 阿部秋生・秋山虔・今井源衛・鈴木日出男 (校注・訳) (1994) 『新編日本古典文学全集 25 源氏物語⑥』, 小学館.
- 市古貞次 (校注・訳) 『新編日本古典文学全集 46 平家物語②』, 小学館.
- 菊地靖彦・木村正中・伊牟田経久 (校注・訳) (1995) 『新編日本古典文学全集 13 土佐日記・蜻蛉日記』, 小学館.
- 阪倉篤義 (校注) (1964) 『日本古典文学大系 78 夜の寝覚』, 岩波書店.
- 鈴木一雄 (校注・訳) (1974) 『日本古典文学全集 19 夜の寝覚』, 小学館.
- 鈴木一雄 (校注・訳) (1996) 『新編日本古典文学全集 28 夜の寝覚』, 小学館.
- 鈴木一雄・石埜敬子 (1984) 『完訳日本の古典第二十五卷夜の寝覚 (一)』, 小学館.
- 鈴木一雄・石埜敬子 (1985) 『完訳日本の古典第二十六卷夜の寝覚 (二)』, 小学館.
- 日本国語大辞典第二版編集委員会・小学館国語辞典編集部 (編) (2000~2002) 『日本国語大辞典』 第2版, 小学館.
- 藤岡忠美・中野幸一・犬養廉・石井文夫 (校注・訳) (1994) 『新編日本古典文学全集 26 和泉式部日記・紫式部日記・更級日記・讃岐典侍日記』, 小学館.
- 三谷栄一・三谷邦明・稲賀敬二 (校注・訳) (2000) 『新編日本古典文学全集 17 落窪物語・堤中納言物語』, 小学館.

### 関連 URL

- |                                |   |
|--------------------------------|---|
| 国立国語研究所『日本語歴史コーパス』             | <a href="https://clrd.ninjal.ac.jp/chj/">https://clrd.ninjal.ac.jp/chj/</a>       |
| 形態素解析器「MeCab」(Ver.0.996)       | <a href="http://taku910.github.io/mecab/">http://taku910.github.io/mecab/</a>     |
| 解析用辞書「中古和文 UniDic」(Ver.202203) | <a href="https://clrd.ninjal.ac.jp/unidic/">https://clrd.ninjal.ac.jp/unidic/</a> |
| プログラミング言語「Python」(Ver.2.7.18)  | <a href="https://www.python.org/">https://www.python.org/</a>                     |