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

平成　３０　年　　６　月　　６　日現在

機関番号：　３２６０１
研究種目：　基盤研究(B)（一般）
研究期間：　2013～2017
課題番号：　２５２８４０８６

研究課題名（和文）平安時代における言語リソースの構築に関する研究

研究課題名（英文）A Study on Construction of Linguistic Resources in Heian Period

研究代表者
近藤　泰弘（KONDO, Yasuhiro）

青山学院大学・文学部・教授

研究者番号：２０１２６０６４

交付決定額（研究期間全体）：（直接経費）　　10,300,000円

研究成果の概要（和文）：平安時代の言語リソースがどういうものであるかについての研究を行った。具体的には、『古今集』や『源氏物語』の語彙をコーパスによって調査し、それらのうちのどの要素が、現代語にまで影響を及ぼす「言語リソース」となっていたかについて考察した。また、言語リソースとなった語がどのように相互に関係しているかについて、機械学習の手法を使って類似度を計算した。それらの研究を元に、国語研究所の日本語歴史コーパスの『源氏物語』をプログラムで処理することで、総合的な言語リソース辞書を作成した。

研究成果の概要（英文）：We conducted research on the linguistic resources of the Heian period. We investigated the vocabulary of "Kokin Wakashui" and "Genji Monogatari" and considered which of them was "linguistic resource". In addition, we calculated word similarity using machine learning method. Based on those studies, we processed the "Genji Monogatari" in the historical corpus of the National Institute for Japanese Language and Linguistics and created a comprehensive language resource dictionary.

研究分野：　日本語学

キーワード：　源氏物語　N-gram　言語リソース　コーパス　XML　構築主義　ジェンダー　機械学習

１．研究開始当初の背景

(1) 現在の言語研究の動向として、歴史言語学においても、社会言語学的な観点の重要性が指摘されるようになってきている。史的語用論という言い方もされるが、言語の体系的な問題、つまり、語彙・文法・意味を個別にとりあげて記述するだけではなく、ある言語体系がどのような社会的価値、イデオロギーによって構築されてきているかについての研究が重要になってきている。本研究を開始するにあたっては、そのような研究動向についての視点が重要であった。

(2) もうひとつ重要な要素は、コーパス言語学の発展による様々な言語コーパスの作成である。特に、本研究にとってなくてはならない古典語コーパスが、国立国語研究所において開発され、日本語歴史コーパスという名称で公開されたことがある。日本語歴史コーパスは、研究代表者がその初代のプロジェクトリーダーであったこともあり、その内容を熟知しているため、本研究に用いるのにもっとも適していた。

２．研究の目的

(1)　本研究の主たる目的は、平安時代語における言語リソースというものの実態を明らかにすることである。そもそも、言語リソースとは、アメリカの言語社会学で用いられている概念であり、ある種の言語的な言い回しや、慣習、スタイルがよってたつ源泉のことである。たとえば、女性語というのは、そのような源泉・資源があり、その枠組みの中から取り出されてくるというイメージである。平安時代語においても、そのような言語リソースが作られていると仮定するならば、そのリソースから次の時代へと引き継がれ、まるで遺伝子のように、現在の日本語に引き継がれてきたものがあるはずである。それが、現代日本語における、たとえば「桜」や「月」や「鶯」といったもののイメージを規定していると思われる。今回の研究では、メタファー・類義語の観点を中心に、平安時代語において、文学の形成における言語リソースを発見することを目的とする。

(2)　第 2 の目標は、コーパス言語学における新しい方法論の開拓である。日本語の古典を対象としたコーパスの作成とその解析については、いまだ、「中納言」等のコーパス解析アプリケーションに頼った方法が一般的である。しかしながら、現在の様々な自然言語処理技術を応用することができれば、さらに多様な研究が可能になってくる。本研究では、研究代表者および連携研究者が開発した、古典語の N-gram による解析手法をさらに発展させて、日本語歴史コーパスの短単位語彙素の N-gram を使った新しい研究手法を開拓することを目的としている。

３．研究の方法

(1)　『古今和歌集』におけるメタファーを採取し、そのパターンを分類した。その後、歌ごとのメタファーを表にして、その他の古典語資料と対比することを試みている。現状はまだこの作業は途中まで進んでいるところであるが、メタファーの観点から見た、言語リソースの形成についての研究手法として、ひとつのスキーマ形成を探ることが可能である。

(2)　短単位語彙素の N-gram による研究手法を用いた。まずは 1 gram の単位を収集し、そのすべてを word2vec と呼ばれる機械学習ソフトウェアで解析し、単語意味の分散処理を行った。その分散表現した結果の類似度を測定し、それぞれの短単位について 10 位までの類似度を持つ短単位を決定した。それぞれの短単位をその語彙素読みの順番に並べ、文脈付きの用例集とした。また、それぞれに、先の類似度による連想語を付載することで、平安時代語の言語リソース辞典としての体裁を整えることができた。

４．研究成果

(1)　まず研究の基礎的部分として、XML 形式のコーパスから辞書形式のデータを作成する研究を主に行った。具体的には次のような手順である。まず、XML 形式 のデータを用意する。

<SUW orthToken="いま" lForm="イマ" lemma="今" lemmaID="2460" kana="イマ" pos="名詞-普通名詞-副詞可能" Form="イマ" pronTo ken="イマ" wType="和" start="20" end="40"orderID="20" /> い ま <SUW orthToken="は" lForm="ハ" lemma="は" lemmaID="29321" kana="ハ" pos="助詞-係助詞" Form="ハ" pronToken="ワ" wType="和" start="40" end="50" orderID="30" />

次にこれを次のような辞書形式に変換する。「03 伊勢,2482,会い見る,アイミル,動詞-一般,文語上一段-マ行,に、手を折りてあひ見し ことをかぞふれば」

そして最後に複数の作品をマージして完成する。このような手続きによって、辞書を作ることができる。今回はこのような研究によって、コーパスから言語リソースを抽出 するための基礎研究を行った。

(2)　研究代表者は、本研究と合わせて国立国語研究所において、通時コーパスプロジェクトに参加して、プロジェクトリーダーとして、日本語歴史コーパスの設計と開発を行った。これによって、研究資料のデジタル化が完成し、本研究の基礎的段階を達成することができた。それを受けて、本研究においては、コーパス処理のためのプログラム開発を行

った。これまでは Perl 言語を用いた自然言語処理の技法を主に使ってきたが、近年の自然言語処理、特に人工知能研究においては、Python 言語を用いた研究が主流となっている。従って、本研究においても、従来の Perl 言語で研究してきたソフトウェアを Python 言語に全面的に書き換えた。また、pandas などのテキスト処理に適したモジュールの採用によって、非常に高速なテキスト処理が可能になった。本研究の重要な成果のひとつである。

(3)　まず最初の段階で、平安初・中期の言語作品から語彙集を作成することを試みた。具体的には次のような形となる。「あ」の冒頭の一部を以下に示す。

あ【吾】[代名詞]「これを見て、「あが仏、何事思ひ(竹取)」「たくみをしたりとも、あの国の人を(竹取)」
あい【合い】[接尾辞-名詞的-一般]「を契る心あらば星あひばかりのかげを見よ(蜻蛉)」
あい【相】[接頭辞]「たり」といへば、あひたてまつる。皇子ののたまはく(竹取)」「はや、この皇子にあひ仕うまつりたまへ」といふ(竹取)」
あいごと【逢い言】[名詞-普通名詞-一般]「酒飲みしければ、もはらあひごともえせで、(伊勢)」
あいだ【間】[名詞-普通名詞-副詞可能]「にて父母あり。かた時の間とて、かの国(竹取)」「え堪へず。このあひだに、ある人の(土佐)」
あいない【あいない】[形容詞-一般]（文語形容詞-ク）「流したてまつると聞くに、あいなしと思ふまでいみじう悲しく(蜻蛉)」「臥して思ひ集むることぞ、あいなきまで多かるを、書き出だし(蜻蛉)」
あいにく【生憎】[形状詞-一般]「ばかりにもあらず、あやにくにあるに、なほ(蜻蛉)」「さにはあらず。あやにくに面嫌ひするほど(蜻蛉)」
あいみる【会い見る】[動詞-一般]（文語上一段-マ行）「に、手を折りてあひ見しことをかぞふれば(伊勢)」「といひて、男、あひ見ては心ひとつ(伊勢)」
あう【会う】[動詞-一般]（文語四段-ハ行）「は、男は女にあふことをす。女(竹取)」「す。女は男にあふことをす。そ(竹取)」
あう【合う】[動詞-非自立可能]（文語四段-ハ行）「をくじり、垣間見、惑ひあへり。さる時より(竹取)」「蓬莱といふらむ山にあふやと、海に(竹取)」
あう【敢う】[動詞-非自立可能]（文語下二段-ハ行）「返しをす。白山にあへば光の失するか(竹取)」「、みのもかさも取りあへで、しとどにぬれ(伊勢)」
あう【あう】[感動詞-一般]「に、齢なども、あうよりにたべければ(蜻蛉)」
あえしらう【あえしらう】[動詞-一般]（文語四段-ハ行）「ぬなめり」などもあへしらひ、硯なる文を(蜻蛉)」「、柚の汁してあへしらひて、まづ出だしたり(蜻蛉)」
あえない【敢え無い】[形容詞-一般]（文語形容詞-ク）「げなきものをば、「あへなし」といひける。(竹取)」「なむ多かりける。さて、あへなかりしすきごとどものそれ(蜻蛉)」
あお【青】[名詞-普通名詞-一般]「なく申すぞ」と、青へとをつきて(竹取)」「、よめりける歌、青海原ふりさけみれば春日(土佐)」
あおい【葵】[名詞-普通名詞-一般]「の実などあるに、葵をかけて、あふひ(蜻蛉)」「、葵をかけて、あふひとか聞けどもよそ(蜻蛉)」
あおい【青い】[形容詞-一般]（文語形容詞-ク）「黒く、松の色は青く、磯の波は(土佐)」「ける人のをなむ、青き苔をきざみて、(伊勢)」
あおうま【青馬】[名詞-普通名詞-一般]「港にあり。今日は白馬を思へど、かひ(土佐)」「とて、世は騒ぐ。白馬やなどいへども、(蜻蛉)」
あおむ【蒼む】[動詞-一般]（文語四段-マ行）「にて、草はところどころ青みわたりにけり。あはれ(蜻蛉)」
あおやぎ【青柳】[名詞-普通名詞-一般]「歌、さざれ波寄するあやをば青柳の影の糸し(土佐)」「ありけるを折りて、あをやぎの糸うちはへて(大和)」
あおり【障泥】[名詞-普通名詞-一般]「ぬ。草のなかにあふりをときしきて、(大和)」
あか【赤】[名詞-普通名詞-一般]「、とまるはただ薄物の赤朽葉を着たるを(蜻蛉)」

(4)　これを「言語リソース」のための土台として、ここからどのような語を最終するかを決定していった。そして、この後、『源氏物語』についても同様なものを作成し、更に、用例を増補、また、word2vec を用いて連想語リストを作成し、語彙集に付加した。同じく「あ」の冒頭の一部を以下に示す。

あ【吾】[代名詞]「、つと抱きて、「あが君、生き出で(4 夕顔)」「抜けば、女、「あが君、あが(7 紅葉賀)」「手を打ちて、「あがおもとにこそおはしまし(22 玉鬘)」「のたまふを、三の宮、「あが大将をや」(37 横笛)」「思たまへわかれず。あが君、とかく押し(39 夕霧)」「心恥づかしくらうたくおぼえて、「あが君、御心(47 総角)」「ほほ笑みぬ。「げに、あが君や、幼(49 宿木)」「おはしませとこそ念じはべれ。あが君は人笑は(50 東屋)」「泣きたまふ。右近、「あが君、かかる御(51 浮舟)」「、乳母なるべし、「あが君や、いづ方(52 蜻蛉)」「ぬにこそあめれ。あが仏、京に(53 手習)」,吾,ブンオウ,継子,伯母,己,物恥ずかしい,吾子,抱く,妹,ゼンギョウ,汝,
ああ【ああ】[感動詞-一般]「もおぼおぼしかりければ、「ああ」と傾きてゐ(34 若菜上)」,

ああ,耄々,旅姿,竜頭,棹差す,唐櫛笥,立蔀,トシカゲ,轟く,黒木,煙たい,
あい【相】[接頭辞]「待ち暮らししを。なほあひ思ふまじきなめり」(2 帚木)」「よ』などやうに、あひ知れる人来とぶらひ(2 帚木)」「、昔人も言ひける。あひ思ひたまへよ。つつむ(3 空蟬)」「こと、いと尊き老僧のあひ知りてはべるに、(4 夕顔)」「をと聞こゆ。人とあひ乗りて簾をだに(9 葵)」「といふが、世にあひはなやかなる若人にて(10 賢木)」「にあり。昔かやうにあひ思し、あはれをも(12 須磨)」「の得意にて、年ごろあひ語らひはべりつれど、(13 明石)」「、その御後はかばかしう相継ぐ人もなくて(18 松風)」「やはらかならむ人をこそあひ思はめと思ふ。(21 少女)」「たまひしを、いかでかあひ語らひ申さむと思ひ(22 玉鬘)」「しかば、重き病をあひ助けてなむ、参り(35 若菜下)」「なり、官位につけてあひ頼む人々、おのづから次々(36 柏木)」「いにしへも何心もなう、あひ思ひかはしたりし世(39 夕霧)」「東と聞こゆなるは、あひ思ひたまひてんや(43 紅梅)」「ぬ心地して、「あひ思せよ。いと心憂く(47 総角)」「承けとり騒ぐめれば、あひあひにたる世の(50 東屋)」「、継母の北の方ことにあひ思はで、兄の(52 蜻蛉)」「小野にはべりつる尼どもあひ訪ひはべらんとてまかり(53 手習)」「、夜半、暁にもあひとぶらはんと思ひたまへ(54 夢浮橋)」,相,揚名,生む,仏法,老僧,瑞歯ぐむ,授ける,出家,下童,罷る,労気,
あいきょう【愛敬】[名詞-普通名詞-一般]「明らかに悟り明かさむこそ愛敬なからめ、などか(2 帚木)」「に、まみ、口つきいと愛敬づき、はなやかなる容貌(3 空蟬)」「笑みたまへる、いとめでたう愛敬づきたまへり。いつ(7 紅葉賀)」「げに生ひなりて、愛敬つき、らうらうじき心ばへいと(8 花宴)」「たまふに、心ばへのらうらうじく愛敬づき、はかなき戯れごとの(9 葵)」「御さまは、いとぞ愛敬づき、いふよしなき(13 明石)」「したまへるが、なかなか愛敬づきて腹立ちなしたまふ(14 澪標)」「顔の何心なきが、愛敬づきにほひたるを、(18 松風)」「たぐへてき」とのたまふ愛敬もこよなし。「禊(20 朝顔)」「語らひて笑ひたまふ。いと愛敬づき、をかしきけさへ(22 玉鬘)」「用意気色などよしあり、愛敬づきたる君なり。(25 蛍)」「て、見るままにいと愛敬づきかをりまさりたまへれ(26 常夏)」「も移り来るやうに、愛敬はにほひ散りて、(28 野分)」「の御けはひ、いと若く愛敬づきたるに、大臣(32 梅枝)」「はただいと切になまめかしう愛敬づきて、見るに(33 藤裏葉)」「ては、似るものなく愛敬づき、なつかしくうつくしきこと(34 若菜上)」「御声たとへむ方なく愛敬づきめでたし。月やうやう(35 若菜下)」「なつかしうなまめき、あてに愛敬づきたまへることの(36 柏木)」「たまへるさまは、いみじう愛敬づきて、にほひやかに(39 夕霧)」「埋れたるさまならず、愛敬づきたまへること、(43 紅梅)」「たまふ、御声あてに愛敬づき、聞かまほしういま(44 竹河)」「おこせて笑ひたる、いと愛敬づきたり。(46 椎本)」「ものから、なつかしげに愛敬づきてもののたまへる(47 総角)」「のいちじるからぬをりだに、愛敬づきらうたきところなどの(49 宿木)」「の少将の君ぞいと愛敬なくおぼえたまふ。こ(50 東屋)」「けり。うち乱れたまへる愛敬よ、まろならば(51 浮舟)」「」とて、笑ひたるまみ愛敬づきたり。声聞く(52 蜻蛉)」「赤めたまへるも、いと愛敬づきうつくしげなり。(53 手習)」,愛敬,ひちちか,しどろもどろ,柔らか,匂いやか,貴,柔柔,物妬み,わららか,子めかしい,あてはか,
あいしゅう【愛執】[名詞-普通名詞-一般]「御契り過ちたまはで、愛執の罪をはるかしきこえ(54 夢浮橋)」,愛執,晴るかす,出家,捉え所,口軽い,満てる,功徳,亡骸,匠,恨み文,本尊,

(5)　このような形で、短単位語彙素の語彙素読み、語彙素代表形、品詞、用例、そして、最後に、連想語を集めた、『源氏物語』語彙集を完成することができた。これはこれまでに全くない新規の形態を持った辞書である。これを用いることによって、平安時代言語の言語リソースのあり方を研究することが可能になったと言える。

(6)　今後、これを辞書化したものを電子テキストとして公開する他、出版の計画も立てている。今後の研究の方向性としては、2 グラム以上の N-gram による複合辞の研究、また N-gram における機械学習の分類によって、平安時代言語のジェンダー性などを研究することが考えられる。これらの方向性を見いだすことができたことも本研究の重要な成果の一つであると考えている。

(7)　最後に、本研究の主要な成果は、すべて、その他の項目にあげたホームページに公開してある。特に、本研究の元となった論文などは英訳も添えて公開することで、今後の国際的な展開も予定している。また、古典語処理プログラムを公開するスペースを設け、機関リポジトリではまだ十分でない古典語処理ソフトウェアの公開も行っている。このように、研究手法の国際化、情報化、公開化の方法を研究することも本研究の重要な目標であったが、MovableType ソフトウェアによって、非常に効率的に公開が可能であることを実証することができた。これも、本研究の成果のひとつとして上げることができる。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 9 件）
① 近藤泰弘、古典語研究におけるコーパス検索と言語の構造との関係、国語語彙史の研究、査読有、Vol.35、2016、pp.1-14、

② 近藤泰弘、平安時代日本語における時間を表す名詞-「時相名詞」の提案と分類-、国語と国文学、査読有、Vol.93、No.5、2016、pp.17-28、
③ 近藤泰弘、歴史コーパスとは何か、日本語学、査読有、Vol.33、No.13、2014、pp.6-15
④ 近藤泰弘、日本語モダリティの史的変遷、ひつじ意味論講座、査読有、Vol.3、2014、pp.119-135、
⑤ 近藤みゆき、コーパスを使った日本文学研究、日本語学、査読有、Vol.33、2014、pp198-206、

〔学会発表〕（計３件）
① 近藤泰弘、言語資源ワークショップ 2017、『源氏物語』コンコーダンスとその応用、2017、国立国語研究所、2017
②近藤泰弘、古典語コーパスからの語彙集作成について、国立国語研究所、2017

〔図書〕（計２件）
① 近藤泰弘・田中牧郎・小木曽智信、ひつじ書房、コーパスと日本語史研究、2015、293

〔その他〕
ホームページ等
http://japanese.gr.jp

６．研究組織
(1)研究代表者
近藤泰弘（KONDO, Yasuhiro）
青山学院大学・文学部・教授
研究者番号：２０１２６０６４

(2)連携研究者
近藤みゆき（KONDO, Miyuki）
実践女子大学・文学部・教授
研究者番号：８０２０５５６７