

共同研究プロジェクト紹介 基幹型：通時コーパスの設計 『日本語歴史コーパス』による平安時代と室町時代の語彙の比較

著者	田中 牧郎
雑誌名	国語研プロジェクトレビュー
巻	6
号	1
ページ	11-20
発行年	2015-06
URL	http://doi.org/10.15084/00000797

『日本語歴史コーパス』による 平安時代と室町時代の語彙の比較

A Lexical Comparison of the Heian Period and the Muromachi Period
Based on the *Corpus of Historical Japanese*

田中 牧郎 (TANAKA Makiro)

1. はじめに

2009年10月に始まった共同研究プロジェクト「通時コーパスの設計」では、日本語の歴史を研究することのできる、通時的な日本語コーパスを設計し、国立国語研究所コーパス開発センターと協力して試作版を作成し、その一部を『日本語歴史コーパス』として順次公開している。設計の基本的考え方は、近藤（2012）にまとめられているので参照してほしい。本稿では、試作版として作成に着手し、一応の完成をみて公開した『日本語歴史コーパス 平安時代編』（国立国語研究所 2014a）、『日本語歴史コーパス 室町時代編 I 狂言』（国立国語研究所 2015）の2つのコーパスを取り上げ、相互の語彙比較を行うことを通じて、日本語史研究における通時コーパスの意義と課題の一端を示す。

2. 『日本語歴史コーパス』の構築

日本語は、文献によってその歴史をたどることができる時代がかなり長い言語であり、言語史研究のために使うことのできる資料の種類や数も相当に豊富である。そのため、日本語史研究における資料研究の蓄積は厚く、この蓄積を踏まえて、コーパスの設計と作成に関する研究を進めていくことになる。

本プロジェクトで行っている、コーパスの設計と作成の研究の基本的な流れは、田中（2014a）に記した。その要点を記すと次の通りである。

- (1) 日本語史研究に限らず、言語史研究は、書き言葉よりも話し言葉の歴史を解明することを第一義とする。そこで、時代別に、話し言葉を反映する度合いの強い資料と、それ以外の重要資料とに分けて資料リストを作成し、まず話し言葉の歴史がたどれる文献を選定し、その後、それ以外の資料を選定することとする。話し言葉を反映する資料リストにあがる文献を個々に吟味し、研究の蓄積が厚く、各時代を代表できる性質の強いものについて電子テキストを入手したり作成したりする。
- (2) 電子テキストは、プレーンテキストに止めず、資料の特性を生かした構造化を行い、タグセットを策定して、マークアップを施し、構造化テキストを作成する（この手順の詳細は、河瀬（2014）参照）。

- (3) 各構造化テキストに対して、時代別の形態素解析辞書を用いて、「短単位」による自動形態素解析を施して単語情報を付与する。その際、誤解析や未知語の有無を目視で確認し、手作業で修正するとともに、解析辞書に正しい情報を追加する（この手順の詳細は、小木曾（2014）参照）。
- (4) 短単位情報が加えられた電子テキストに対して、「長単位」による単語情報を付与する（長単位については、富士池（2015）を参照）。
- (5) 短単位および長単位の情報によって検索できる、Web 検索ツール『中納言』に、単語情報付きデータを格納して、公開する（公開 URL は、<https://maro.ninjal.ac.jp/>）。

(1) に関しては、これまでに公開したコーパスに、『日本語歴史コーパス 平安時代編』と『日本語歴史コーパス 室町時代編 I 狂言』があるが¹、それぞれ、和文文学作品（底本は小学館刊行の『新編日本古典文学全集』、「古今和歌集」「枕草子」「源氏物語」など 14 作品）と虎明本狂言集（底本は清文堂刊行の『大蔵虎明能狂言集翻刻註解』（大塚光信編，清文堂出版，2006））を対象にしており、いずれも、平安時代と室町時代の話し言葉を反映した資料として、第一に取り上げられるべきものである。

(2)～(5) は、資料を読み込み、電子的に翻刻し、文章を読解し、単語を認定し、それらのデータをコンピュータに格納していく、長期間にわたって持続する、文献学的言語研究の基本となる作業であり、コーパス作成過程の中心を占めている。平安時代編はこれらの作業のすべてが完了しているが²、室町時代編は(4)の長単位の情報付与が未了である。そのため、(5)の『中納言』によって公開されているのは、平安時代編は、短単位と長単位、室町時代編は、短単位のみとなっている。室町時代編の長単位は、後日公開する予定である³。

公開されているコーパスデータを、短単位（記号を除く）で集計すると、『日本語歴史コーパス 平安時代編』が約 74 万語、『日本語歴史コーパス 室町時代編 I 狂言』が約 23 万 5 千語である。このデータに対して、いくつかの語彙調査を行ってみよう。

3. 語種構成比率の変化

上記の 2 つのコーパスにおける短単位データから、付属語と固有名詞を除外すると、『日本語歴史コーパス 平安時代編』は、延べ語数 418,279 語、異なり語数 8,861 語であり、『日

¹ ほかに近代語のコーパスとして、『明六雑誌コーパス』（国立国語研究所 2012）、『国民之友コーパス』（国立国語研究所 2014b）を公開しているが、現段階では『中納言』による検索に対応していないなど、仕様が異なるところがあるので、本稿では扱わない。近代語のコーパスについても、近い将来、『日本語歴史コーパス』に組み入れる計画があり、その際に仕様を統一することになるだろう。なお、近代語のコーパスについては、2009 年度から 2012 年度までに実施された、独創・発展型共同研究プロジェクト「近代語コーパス設計のための文献言語研究」（プロジェクトリーダー：田中牧郎）の成果を、本プロジェクトが受け継いでおり、それ以前の 1980 年代に始まった「日本大語誌」計画を受け継ぐ部分もある。それら、近代語のコーパスの設計と構築については、田中（2014b）に記したので、参照してほしい。

² 平安時代編には、和文作品を増補する予定がある。また、将来的には、平安時代の漢文系資料を組み入れる必要もあるが、これについての計画は具体化していない。

³ 『日本語歴史コーパス 室町時代編 I 狂言』の作成から公開に至るまでの作業上の問題点は、市村・渡辺ほか（2015）にまとめられている。

本語歴史コーパス 室町時代編 I 狂言』は、延べ語数 130,800 語、異なり語数 8,540 語となる⁴。以下、これらをそれぞれ、「平安和文」「室町狂言」の語彙と扱って分析していく。延べ語数では、平安和文は室町狂言の 3 倍以上になるが、異なり語数では、両者に大きな差がない。このことは、平安和文は、少ない種類の語彙で書かれているのに対して、室町狂言は、多くの種類の語彙で書かれていることを示している。

平安和文と室町狂言の語彙について、「和語」「漢語」「外来語」「混種語」の語種構成比率を調査すると、図 1（延べ語数）、図 2（異なり語数）のようになる⁵。この語種情報はコーパスに付与されているものを用いているが、そこで「外来語」とされるもののほとんどは、梵語由来の語（「菩提」「旦那」など）である。

図 1、図 2 いずれにおいても、平安和文から室町狂言への最も目立つ変化は、漢語の増加である。漢語は、延べ語数で約 3 倍、異なり語数で約 2 倍に増加している。また、総数は少ないが、混種語も延べ語数で約 8 倍に増加している。異なり語数に比べて延べ語数の伸びが

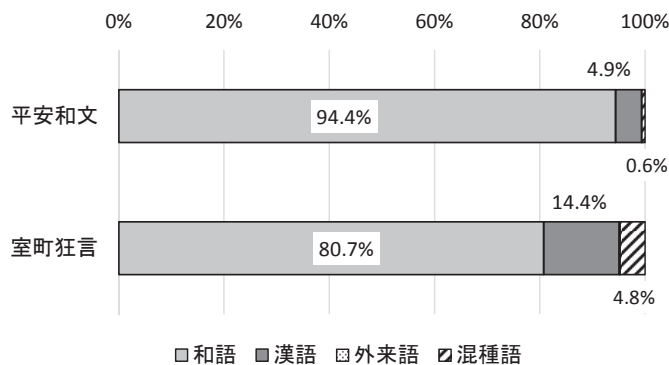


図 1 語種構成比率（延べ語数）

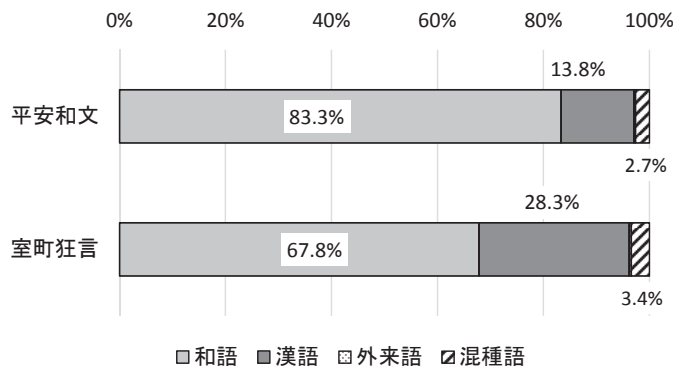


図 2 語種構成比率（異なり語数）

⁴ 室町時代編の数値は公開前の暫定版に基づいたもので、公開版では変更になる可能性がある。

⁵ 語種が「不明」とされる、平安和文の 3 語、室町狂言の 24 語（いずれも延べ語数）を除く集計。

著しいのは、頻繁に使われる漢語や混種語が増えているということである。混種語のほとんどは漢語と和語が組み合わさったものであるので、総じて頻繁に使われる漢語が増えているということである。次節では、そこに注目して、漢語にどのような変化があったのかを、具体的に見ていこう。

4. 高頻度語彙の変化

4.1 高頻度語彙における語種構成比率

語彙調査において高頻度を示す語彙は、色々な場面で繰り返し使われる基本的な語彙であると考えられる。語を頻度順に並べて、上位の語から使用率（総語数の中に占める比率）を累積した「累積使用率」（カバー率）が80%に達するまでの語彙の範囲を、高頻度語彙と扱うことにすると、平安和文の高頻度語彙は、度数88（730位）までの741語、室町狂言の高頻度語彙は、度数18（990位）までの1,037語となる。高頻度語彙の語数にも、少ない種類の語彙で書く傾向が強い平安和文と、多くの種類の語彙で書く傾向の強い室町狂言の違いが見えているとすることができる。

図3は、その高頻度語彙における語種構成比率（異なり語数）を、平安和文と室町狂言とで比較できるように図示したものである。

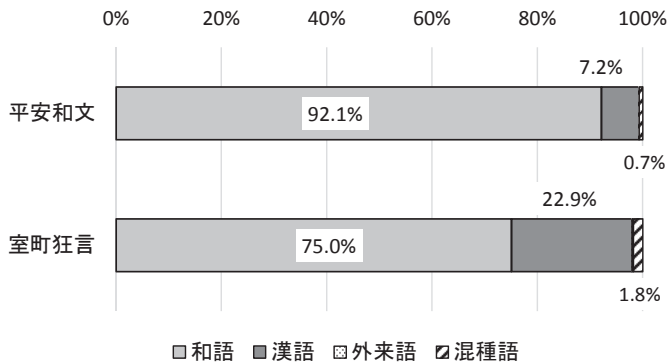


図3 高頻度語彙における語種構成比率（異なり語数）

図3によると、室町狂言は平安和文に比べて、漢語の比率が大幅に増加し3倍以上となっていることが見て取れる。また、全体の比率は非常に小さいが、混種語も2~3倍に増加していることがわかる。以下、増加が著しく量も多い漢語について、詳しく見ていこう。

4.2 高頻度の漢語の品詞と意味の変容

図4は、高頻度語彙に属する漢語における品詞構成比率を、平安和文と室町狂言とで比較して示したものである。品詞情報は、『日本語歴史コーパス』に付与されているものをもとに、大きな分類にまとめ直して集計した。例えば、「名詞-普通名詞-形状詞可能」などとあるも

のは、コーパスの用例に名詞用法が多いものは「名詞」、形状詞用法が多いものは「形状詞」とした。図4で「助動詞」とあるのは「様(よう)」で、これは『日本語歴史コーパス』では「名詞-普通名詞-形状詞可能」とされるが、学校文法では助動詞「様なり」とされ、他の漢語名詞とは質を異にするので、ここでも「助動詞」と扱った。

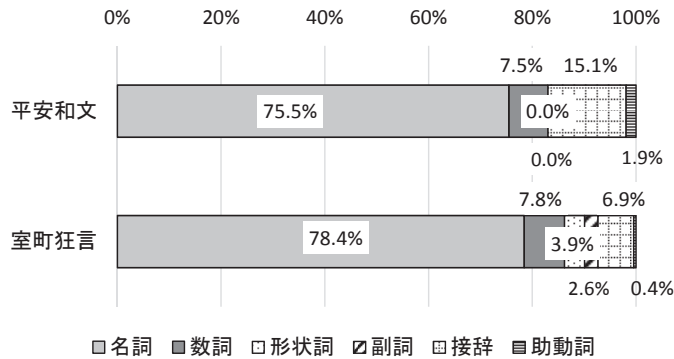


図4 高頻度の漢語における品詞構成比率

図4によれば、平安和文においても室町狂言においても名詞が70~80%を占め、数詞が7~8%占めているところは不変だが、平安和文に多かった接辞が室町狂言で、15.1%から6.9%へと大幅に減少し、代わって平安和文にはなかった、形状詞と副詞が登場しているという変化をとらえることができる。その変化した接辞、形状詞、副詞を一覧にすると、表1のようになる。

表1 高頻度の漢語における接辞・形状詞・副詞

	平安和文	室町狂言
接辞	故-, 左-, 大-, 中-, -位, -院, -卿, -人	御-, 大-, 不-, -御, -国, -山, -寺, -者, -衆, -天, -人, -匹, -邊, -坊, -本, -様
形状詞		一段, 有徳, 結構, 散々, 殊勝, 大事, 沢山, 利根, 聊爾
副詞		再々, 自然, 随分, 是非, 早々, 内々

まず接辞を見ると、平安和文では、漢語の中で占める比率が比較的高いといっても、その大部分は「故大納言」「左大臣」「五位」など、職階を表す語の一部になっているものであって、その意味には限定がある。一方、室町狂言では、漢語の中で占める比率は低くなっている、職階を表す語の一部という限定はなくなり、「女房衆」「案内者」など人を表す語や、「御存じ」「不案内」のような抽象概念に付く語なども多く、多様な意味に広がっている。平安和文になかった形状詞や副詞の個々の語を見ても、室町狂言では、多様な意味の語に広がっ

ていることがわかる。高頻度を示す基本的な語彙に組み込まれた漢語が、多様な品詞や意味に広がってきているのである。

次に、平安和文でも室町狂言でも最も多数を占める名詞の内訳を見よう。図5は、高頻度語彙に属する漢語の名詞を、『分類語彙表 増補改訂版』（国立国語研究所 2005）の分類番号に依拠して5つの意味に分類し、その比率を示したものである。なお、『分類語彙表 増補改訂版』に掲載されていない語は、掲載されている語で最も近い意味の語と考えられるものの番号により、多義語の場合は、最もよく用いられる意味の番号によった。

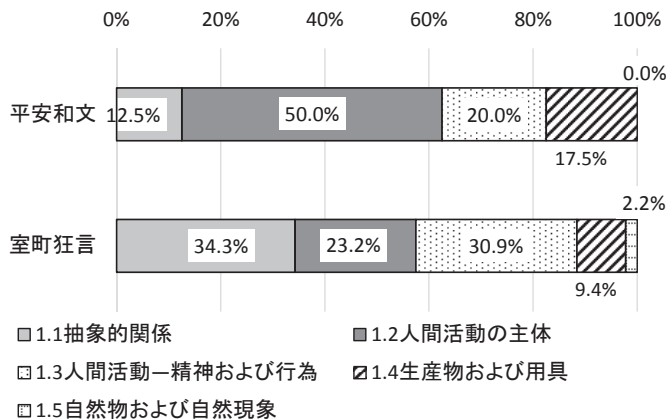


図5 高頻度の漢語名詞の意味構成

図5を見ると、高頻度語彙の漢語名詞の意味は、次のように変化している。すなわち、平安和文で最大の比率を示していた「1.2 人間活動の主体」は、室町狂言では半減し、3番目に多かった「1.4 生産物および用具」も、同じく半減している。反対に、平安和文で比率の大きくなかった「1.1 抽象的關係」、それから「1.3 人間活動—精神および行為」が、室町狂言では、それぞれ3倍近く、1.5倍以上に増加している。つまり、平安和文では、人や物といった具体物を指す語が多くを占めていたのが、室町狂言では、関係や活動など抽象概念を指す語が多くを占めるように、高頻度の漢語の意味のありようが変容しているのである。

4.3 高頻度漢語の意味の広がり

高頻度の漢語名詞が、具体物を指し示す語から抽象概念を表す語へと変化したことについて、語彙リストを作成して比較することで、詳しく見よう。表2は、平安和文と室町狂言のそれぞれについて、高頻度の漢語名詞を、分類語彙表の5分類ごとに一覧にしたものである。

表2によると、比率では減少しているように見えた、「1.2 人間活動の主体」「1.4 生産物および用具」も、語数では、大幅に増加している。例えば、1.2では、平安和文では、皇族や貴族の職階を指す語が多いが、室町狂言では、様々な身分や職業を指す語があり、1.4では、平安和文では、寝殿やそこに置かれる調度を指す語が多いが、室町狂言では、楽器、飲食物、

菓など多様な物を指す語が見られ、それぞれ、意味の範囲は大幅に広がっているのである。

比率において増加が目立った、「1.1 抽象的關係」「1.3 人間活動—精神および行為」においては、室町狂言で非常に多くの漢語があがっており、その意味の範囲は『分類語彙表 増補改訂版』が設定している「中項目」のほとんどすべてに及び、多様な意味に漢語が広がっていることが確かめられる。

表 2 高頻度の漢語名詞

	平安和文	室町狂言
1.1 抽象的關係	例, 宿世, 気色, 頓, 日	分, 儀, 沙汰, 別, 真実, 不思議, 是非, 主, 一段, 縁, 証拠, 在京, 留守, 面目, 様子, 次第, 体, 子細, 奇特, 無理, 用, 路次, 道断, 同道, 下向, 打擲, 普段, 時宜, 時分, 拍子, 毎年, 後生, 代, 吉日, 今日, 今夜, 今度, 先度, 当年, 最前, 舞台, 方々, 地, 座, 洛中, 遠国, 満足, 余, 両, 兩人, 番, 年, 月, 尺, 寸, 度, 行, 首, 字, 騎, 杯, 面
1.2 人間活動の主体	院, 三位, 中宮, 女房, 春宮, 法師, 僧都, 右近, 納言, 女御, 命婦, 宰相, 大将, 中将, 少将, 侍従, 京, 衛門, 兵部, 弁	人間, 衆, 愚僧, 面々, 神, 明神, 地藏, 精, 餓鬼, 坊主, 亭主, 女房, 三郎, 次郎, 太郎, 舍弟, 王, 大名, 下人, 罪人, 役, 新発意, 僧, 法師, 上人, 座頭, 大臣, 博労, 百姓, 仏師, 冠者, 師匠, 弟子, 奏者, 在所, 京, 日, 天下, 地獄, 極楽, 六道, 西門
1.3 人間活動—精神および行為	本意, 用意, 題, 御覧, 経, 消息, 絵, 対面	気, 機嫌, 臆病, 迷惑, 御免, 一声, 念, 堪忍, 所望, 恩, 念仏, 稽古, 御意, 思案, 才覚, 異見, 不審, 用心, 合点, 分別, 算用, 夢想, 慮外, 道理, 法, 約束, 見物, 言語, 談合, 経, 連歌, 秀句, 果報, 福, 出家, 通夜, 芸, 遊山, 礼, 座禅, 富貴, 公事, 推参, 案内, 奉公, 勝負, 訴訟, 成敗, 秘蔵, 無用, 年貢, 扶持, 商売, 布施, 普請, 重宝
1.4 生産物および用具	装束, 対, 殿上, 障子, 格子, 几帳, 屏風	代物, 棒, 烏帽子, 数珠, 昆布, 酒, 茶, 膏薬, 新座, 床机, 道具, 包丁, 羯鼓, 太鼓, 尺八, 人形, 制札
1.5 自然物および自然現象		衆生, 柑子, 雁, 息災

5. 日本語史研究における通時コーパスの意義と課題

語種構成比率の時代による変化を扱った先行研究に前田（1984）がある。そこでは、平安和文、鎌倉時代の軍記、室町時代の抄物などの語彙調査結果が示され、「全体としては時代とともに漢語・混種語の使用率が高まっていく傾向がある」とされる一方で、「文体の違いをこえて、時代的な変遷として考えるのは困難である」「現段階では、語種の量的構造の研究は、語彙というよりは文体を定めるものとして役立っているとしか言えないように思われ

るのである」と述べられている。鎌倉時代の軍記は和漢混淆文であり、室町時代の抄物は原拠である漢文が踏まえていることから、漢文訓読体の影響で漢語の比率が高くなっていて、それは時代差ではないという解釈である。その点、本稿で扱った、平安和文と室町狂言とは、ともに両時代の話し言葉を色濃く反映しており、漢文訓読体の影響は考えにくいいため、漢語の増加は、文体差を要因とするよりも、時代差を要因とすると解釈できる。話し言葉を反映する度合いの強い、漢文訓読体の影響のない資料を、通時的に比較できる設計にしたことによって、漢語の増加という語彙史上の事実が、明確にとらえられたわけである。

一方で、平安和文は貴族層に限定された仮名文学であり、室町狂言は庶民も含む幅広い層の演劇であるという、ジャンルの違いは厳然と存在している。本稿で明らかになった、平安和文が少ない種類の語彙でまかなわれているのに対して、室町狂言が多くの種類の語彙を用いているのは、それぞれのジャンルが扱う世界の狭さ広さに基づく差異だと考えられようし、高頻度の漢語における人や物を指す語の意味が、平安和文ではかなり限定的であるのに対して、室町狂言では多様な範囲に及んでいるのも、それぞれのジャンルが描く事物の狭さ広さの違いに関わっていよう。このジャンル差を解消しようと、平安時代に庶民の話し言葉が反映した演劇ジャンルの資料を求めても、それは存在しない。他方、室町時代には貴族層の仮名文学に連なるジャンルとして御伽草子があり、これをコーパス化することで、ジャンルに配慮した通時的な比較を行うことは望まれるだろう。ただし、御伽草子は、室町時代の話し言葉を反映する性格は弱く、鎌倉時代の説話や軍記の流れも汲んでおり、平安和文とは質の異なるところも多く、時代差以外の差異をすべて解消することは、そもそも不可能なことである。このような、文体、ジャンル、話し言葉性といった、様々な観点で、資料の性質を吟味して、通時的なコーパスを設計し、その設計に沿って作成したコーパスの分析を重ねていくことは、コーパスに基づく日本語史研究の意義にほかならないだろう。

また、本稿では、単に漢語が増えたということだけではなく、高頻度を示す基本的な語彙の中に漢語が大幅に増えていることも示した。こうした語彙全体を把握した上で、その内部構成を頻度によって分類してとらえる研究は、従来は非常に手間がかかるものであったが、単語情報が付与されたコーパスができたことによって比較的容易に行うことができるようになった。さらに、基本的な語彙に位置付く漢語の意味について、『分類語彙表 増補改訂版』の意味に基づいて把握して比較することも行った。語の頻度と意味とを関連付けて分析することで、語彙史上の重要な事実の数々が解明されることが期待できる。しかしながら、『日本語歴史コーパス』に『分類語彙表 増補改訂版』の分類番号が付与されているわけではなく、今回は、高頻度語彙における漢語のリストに対して、筆者が手作業で分類番号を付与したデータによっている。解析辞書に、あらかじめその番号を格納しておき、その辞書で解析したコーパスの単語情報にこれが含まれるようにすることなどが想定できるが、その実現のためには、語の意味の研究をコーパスの設計の中に位置付けて研究していく必要があるが、現段階では十分に進んでいない。これは、宮島ほか（2014）のように語彙表レベルでは実現できているものであり、これをコーパスに取り込んでいく研究が望まれる。

6. おわりに

以上、本稿では、平安和文と室町狂言との語彙比較を通して、通時コーパスによる日本語史研究の意義と課題の一端を述べてきた。2015年度末の本プロジェクト終了までに、『日本語歴史コーパス 鎌倉時代編』を公開し、平安時代と室町時代を繋ぐ計画があるほか、2016年度から始まる後継プロジェクトでは、奈良時代、江戸時代、明治・大正時代のコーパスの作成と公開を行うことも立案中であり、日本語の通史をとらえることのできるコーパスが形をなすのは遠い将来のことではなくなっている。その一方で、そのコーパスをどのようなものにしていけば日本語史研究に真に有用であるのか、コーパスをどのように活用すべきなのかという研究を、いっそう活発に行うことも、強く求められている。

なお、本共同研究プロジェクトの成果をまとめた雑誌特集号に、近藤・田中ほか(2014)が、論文集に、近藤・田中・小木曾(2015 予定)があるので、参照してほしい。

●参考文献●

- 富士池優美(2015 予定)『『日本語歴史コーパス 平安時代編』の形態論情報』近藤泰弘・田中牧郎・小木曾智信(編)『コーパスと日本語史研究』東京：ひつじ書房。
- 市村太郎・渡辺由貴・鴻野知暁・河瀬彰宏・小林正行・山田里奈・堀川千晶・村山実和子・小木曾智信・田中牧郎(2015)『『虎明本狂言集』コーパスの公開』『日本語学会 2015 年度春季大会予稿集』, 167-172。
- 河瀬彰宏(2014)『『日本語歴史コーパス』の設計を支えるマークアップとはなにかー基礎データはどのようにして作られるのかー』『日本語学』33(14):68-82. 東京：明治書院。
- 国立国語研究所(2005)『分類語彙表 増補改訂版』(国立国語研究所資料集 14). 東京：大日本図書。
- 国立国語研究所(2012)『明六雑誌コーパス』国立国語研究所コーパス開発センター Web サイト。
http://pj.ninjal.ac.jp/corpus_center/cmj/
- 国立国語研究所(2014a)『日本語歴史コーパス 平安時代編』国立国語研究所コーパス開発センター Web サイト。 http://pj.ninjal.ac.jp/corpus_center/chj/
- 国立国語研究所(2014b)『国民之友コーパス』国立国語研究所コーパス開発センター Web サイト。
http://pj.ninjal.ac.jp/corpus_center/cmj/
- 国立国語研究所(2015)『日本語歴史コーパス 室町時代編 I 狂言』国立国語研究所コーパス開発センター Web サイト。 http://pj.ninjal.ac.jp/corpus_center/chj/
- 近藤泰弘(2012)「日本語通時コーパスの設計について」『国語研プロジェクトレビュー』3(2): 84-92. 国立国語研究所。
- 近藤泰弘・田中牧郎ほか(編)(2014)『日本語学 臨時増刊 特集 日本語史研究と歴史コーパス』33(14). 東京：明治書院。
- 近藤泰弘・田中牧郎・小木曾智信(編)(2015 予定)『コーパスと日本語史研究』東京：ひつじ書房。
- 前田富祺(1984)「語種構造の漸移相」『日本語学』3(9): 28-39. 東京：明治書院。
- 宮島達夫・鈴木泰・石井久雄・安部清哉(2014)『日本古典対照分類語彙表』東京：笠間書院。
- 小木曾智信(2014)「歴史コーパスにおける形態素解析と辞書整備」『日本語学』33(14): 83-95. 東京：明治書院。
- 田中牧郎(2014a)『『日本語歴史コーパス』の構築』『日本語学』33(14): 56-67. 東京：明治書院。
- 田中牧郎(2014b)「歴史コーパス」前川喜久雄(監), 山崎誠(編)『講座日本語コーパス 2 書き言葉コーパス』, 116-138. 東京：朝倉書店。

《要旨》 本論文では、『日本語歴史コーパス』を用いて、平安和文と室町狂言の語彙調査を行った。その結果、全体として漢語と混種語が大きく増加していることと、高頻度の基本的な語彙においてもそれらの語種が大きく増加していることがわかった。また、漢語はその数が増加しただけではなく、その意味の範囲も拡大させていた。この変化は、時代差だけでなく、ジャンル差によるところもあると考えられる。以上の結果をもとに、コーパスによる日本語史研究の意義と課題について議論した。

Abstract: In this paper, I investigate lexical changes from Heian Wabun to Muromachi Kyogen using the *Corpus of historical Japanese*. The results show that the number of Sino-Japanese words and hybrid words increased significantly overall. The results also show that such words increased significantly in the high-frequency basic vocabulary, and that their semantic range expanded. These changes are due not only to the time difference but also to the genre difference. Based on these results, I discuss the significance of diachronic corpora for historical studies of Japanese and also some of the problems that arise.

田中 牧郎 (たなか・まきろう)

明治大学国際日本学部教授、国立国語研究所言語資源研究系客員教授。博士(学術)(東京工業大学)。昭和女子大学講師、国立国語研究所研究員、同主任研究員、同グループ長、同准教授等を経て、2014年4月より現職。

主な著書・論文：『雑誌『太陽』による確立期現代語の研究—『太陽コーパス』研究論文集—』(国立国語研究所報告122, 博文館新社, 2005), 『病院の言葉を分かりやすく—工夫の提案—』(国立国語研究所「病院の言葉」委員会編, 勁草書房, 2009), 『外来語研究の新展開』(共編著, おうふう, 2012), 『そうだったんだ! 日本語 近代書き言葉はこうしてできた』(岩波書店, 2013), 『『日本語歴史コーパス』の構築』(『日本語学』33(14), 明治書院, 2014)。

社会活動：日本語学会評議員, 日本語学会『日本語学大辞典』編集委員主任, 明治書院『日本語学』編集委員, 日本医学学会用語管理委員会委員。

基幹型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

(明治大学 国際日本学部 教授 / 国立国語研究所 言語資源研究系 客員教授)

プロジェクトの概要

平安時代の和文、鎌倉時代の説話、室町時代の狂言、江戸時代の洒落本、明治・大正時代の雑誌などを対象とした、「通時コーパス」のモデルを試作することを通して、通時コーパス設計における重要問題について研究する。具体的には、①どのような観点で資料を選定するか、②古典本文をどのように電子化しどのような情報(異文、原文表記、異体字、引用、文体など)を付与するか、③各時代の文法・語彙に対応した形態素解析をどのように行うか、などについて、コーパスの試作と活用を通して明らかになる諸問題を解決するための検討を行うことで、通時コーパス構築のための理論的な検討を十分に重ねていく。試作したコーパスを共同研究者が活用してどのような日本語史研究が実施できるかを持ち寄って発表する研究会を重ね、コーパスを評価しつつ、本格的に構築するコーパスの設計を固めていく。