

アパブランシャ語コーパスを用いた韻律による年代推定 Estimating Chronology of Apabhramśa Texts from Meters

山畑 倫志

Tomoyuki Yamahata

北海道大学大学院 文学研究科, 北海道札幌市北区北 7 条西 5 丁目

Hokkaido University Graduate School of Letters, Kita 10, Nishi 7, Kita-ku, Sapporo

あらまし: アパブランシャ語はインド語派の中期インド語に属する言語である。中期インド語から近代インド諸語へ至る過程は未だ不明瞭であり、その間アパブランシャ語はその解明のために重要である。しかし長期間広い地域で使用されたため、言語的に多様な変異を含み、言語事象のみでの分析は困難である。そこで文献がすべて韻文であることを利用し、DB 化した作品群をもとに各詩節と韻律の発展過程を分析し、一定の分類基準を見出す。

Summary: Apabhramśa language is a one of Middle Indic Languages. Though Apabhramśa is important for study of Historical Indic Linguistics, there are plenty of different words, morphological forms and meters. This paper applied the Decision Tree classifier and the Naive Bayes classifier to Apabhramśa corpus for discovering the conditions of these differences. Therefore, we found a new clue for classification of Apabhramśa language.

キーワード: データベース, コーパス, アパブランシャ語, 文書分類, 機械学習

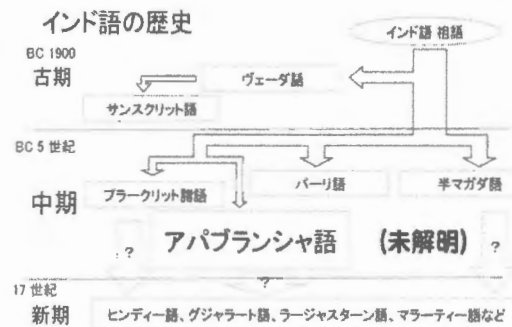
Keywords: Database, Corpus, Apabhramśa, Document Classification, Machine Learning

1. アパブランシャ語の概要

本稿はインド・ヨーロッパ語に属するアパブランシャ語についてコーパスと韻律情報を用いて年代推定を目的とする。まずアパブランシャ語の歴史的な位置づけ、使用状況および言語特徴について概説する。

インド語 (Indic, Indo-Aryan) は紀元前 1900 年ごろからその資料が入手可能な言語であり、またその後も絶え間なく言語資料を作り出し、伝えてきた言語である。そのため言語がどのように変化するかを探る上で重要な言語である。本発表でとりあげるアパブランシャ語 (Apabhramśa) はインド語の歴史の中では中期インド語の最新層に分類される言語である。(図表 1 参照)

中期インド語には多くの言語が含まれ、多様な文献が多く残されている。しかし、それらの歴史的および地域的な整理についてはそれ以前の古期インド語や、ヒンディー語をはじめとした新期インド語と比べ、あまりすすんでいないのが現状である。特に現在インド各地で使用されている新期インド諸語とどのような関係があるのかがいまだ不明なままである。その原因として各言語における言語特徴が錯綜しており、それぞれの地域

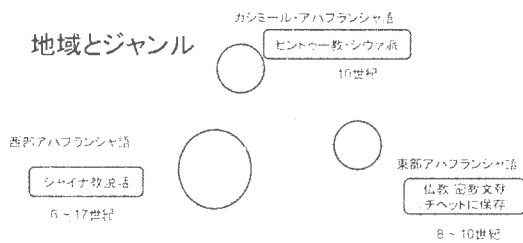


図表 1 インド語の歴史

における新期インド語と簡単に関連づけられないことがある。そのような状況の中で中期インド語の中でもっとも新しいとされるアパブランシャ語は新期インド語への展開を考える上で重要な言語である。

アパブランシャ語は紀元六世紀ごろから十七世紀頃までインド西部を中心に北インドの各地で用いられた言語である。言語名である「アパブランシャ」はサンスクリット語で「墮落した」を意味することから、元来は一地方語であり、サンスクリット語およびマハーラーシュトラ語やシューラセーナ語といった早くから文章語と

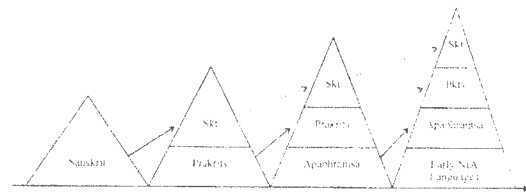
して高い地位にあった言語とは異なった扱いを受けていたことが推測できる。それが具体的にどのような経緯によって文章語となったかは未だ定かではないが、徐宗教詩や宗教説話の分野での使用が徐々に広がり、十二世紀に至るとグジャラート地方のジャイナ教徒ヘーマチャンドラ(Hemacandra)により文法や詩作法に関する作品が著され、一定程度の規範化がなされるまでになった。その後インド西部においてジャイナ教徒が中心となって類似した形式で多くの説話作品が作られた。それに対しヘーマチャンドラ以前から相当数の作品がアハブランシャ語で作られている。そのため、言語特徴上多くの差異を含むテキスト群が残っている。それらの言語特徴はその作成地域にもとづき図表2のように分類可能である。



図表2 アハブランシ語の地域とジャンル

この分類は新期インド語との関連を示唆するため、同様の分類から各地のインド語との関連づける試みがなされてきた。¹しかし、これまでの研究により、そのような地域的な特徴と結びつけることは難しいことがわかってきた。むしろある特定地域の言語が一時的に北インドの広い地域で使われた時代が存在したと考えるのが妥当である。実際に使用されてきた時期の長さ、制作された作品の多さ、また文学スタイルの伝統が類似を考えるとインド西部地域の言語であった可能性が高い。そうした場合、アハブランシャ語に含まれる差異を地域性とは別の形で説明する必要が出てくる。本稿ではそれを文章語としての定型化の過程の中で生じたものと捉え、ヘーマチャンドラによる規範化以前に各地での流行が終わってしまったため、インド西部以外の地域では規範化の途上の特徴を有していると解釈する。Bubenik(1998)はアハブランシャ語が他の諸言語とともに文章語として採用されていく過程を図表3のように示している。

次にアハブランシャ語の言語特徴であるが、インド語の歴史は大まかに言ってサンスクリット語のように複雑な語尾変化で単語間の関係を示す方法から、語尾



図表3 インド文章語の展開 (Bubenik 1998)

変
化を
単
純化し語順と後置詞によって文法機能を表すという方法に大きく変化した。アハブランシャ語はその過程のちょうど中間に位置するため、どちらの特徴も有している。Bubenik(1998)はインド語の文法構造における歴史的变化という枠組みの中でアハブランシャ語を分析している。そこで示される特徴は次の四点である。

I 形態的に示される格語尾が減少し、また単純化されているため、格標示のための形式がほとんどの語幹に共通している。そのため後置詞とみなしうる語が登場してきている。

II 未完了、アオリスト、完了を一語の活用で表す定動詞の形式が完全に消滅する。それにより過去を表現する形式が過去受動分詞のみになる。

III 本来は「立つ」や「ある」を意味する動詞が助動詞的な役割を持ち、それが現在分詞や過去分詞と共に使われて進行や完了などのアスペクトを表している。

IV 過去の表現が過去受動分詞のみになった結果、完了表現における能格構造の成立が見られる。つまり、自動詞主語と他動詞目的語が同じ形式を取り、他動詞主語が独自の形を取っている。

上記をまとめると一つの文法的機能を複数の語で示すいわゆる孤立語的な傾向を示している。それに加えてIVでは能格構造の出現が見られる。屈折型から孤立型へ、そして主格構造から能格構造という大きな変化の途上にあるのである。このような中間的な性格を持つ言語を分析し、他の典型的な構造を持つインド語以外の諸言語との比較により、言語構造一般の理解にとって得るところの多いものとなる。上に挙げた諸特徴は全て近代インド諸語と共通する。しかし、動詞の活用の体系が存続しているなど、古い屈折的な性格も持つ。インド語史の観点からみれば、まさに中間的特徴を有する言語である。

¹ Tagare (1948)

2. アパブランシャ語コーパス作成の試み

本稿で用いたコーパスは以下のテキストから構成されている。便宜的に図表 2 で示した地域区分ごとに示す。また制作年代を付しているが、インドの古典文献一般の特徴として作品の年代がはっきりとわからないことが多い。この中で年代推定の確度が高いのは当時の碑文から確認が取れるヘーマチャンドラのみであり、ほかは相対的な年代推定に基づくものがほとんどである。

西部アパブランシャ語

『ヴィクラム王とウルヴァシー』(部分)

カーリダーサ作、6 世紀?

古典サンスクリット文学の代表的な詩人であるカーリダーサの手になる戯曲作品。主人公のヴィクラム王が悲しみで我を忘れた際の台詞がアパブランシャ語である。アパブランシャ語の最古の用例となるが、その部分が欠ける写本の系統も存在するため、真偽について意見が割れている。

『パドマの行跡 (*Paumacariu*)』(PC)

スヴァヤンブー作、9~10 世紀

インドの代表的な叙事詩である『ラーマヤナ』はバラモンあるいはヒンドゥーの価値観で書かれた物語であるが、それをジャイナ教の教義に合わせて再構成した作品。De Clercq (2003) により電子テキストが提供されているため、それを用いてコーパスを作成。

『ハリシェーナの行跡 (*Harisenacariu*)』10 世紀以降?

ジャイナ教説話では六十三偉人という伝説上の人物の伝記を大枠としてその中に種々雑多な題材を含ませる形式の作品が多い。この作品もその偉人の一人の伝記の形式をとった作品である。

『サナトクマーラの行跡 (*Sanatkumāracarita*)』(部分)

ヘーマチャンドラ、12 世紀

アパブランシャ語の規範化を進めたヘーマチャンドラによる作品。全体としてはサンスクリット語の作品だが、アパブランシャ語部分も存在する。

東部アパブランシャ語

『ドーハーの宝庫 (*Dohakoṣa*)』

カーンハ作、7~12 世紀?

『ドーハーの宝庫』

サラハ作、11~12 世紀?

これら東部アパブランシャ語作品はすべて仏教徒によるものであり、密教の教義を伝える韻文となっている。

カシミール・アパブランシャ語

『タントラ綱要 (*Tantrasāra*)』(部分)

アビナヴァグプタ作、10~11 世紀

同著者の宗教理論書である『タントラ・アーローカ』の綱要であり、各章の末尾がアパブランシャ語の韻文となっている。

『聖典騒動 (*Āgamadambara*)』(部分)

ジャヤンタバッタ作、9~10 世紀?

戯曲作品。正統バラモンを自認する主人公が他の宗教宗派を攻撃する作品。登場人物のうちニーランバラという宗派の行者がアパブランシャ語の韻文を歌いながら登場する。

以上 8 種のテキストを入力し、単語単位に分割し、コーパスとして利用している。インド古典語の言語処理においてしばしば問題となる連声 (*sandhi*)²の問題であるが、アパブランシャ語では原則連声の現象は消失しているため、その問題は生じない。また複数の名詞を複合語として圧縮し句や文に近い機能をもたせる複合語生成もインド古典語に特有の現象だが、作成したコーパスでは標を付与した上で一語として数えている。

| 地域 | トークン | タイプ | タイトル | トークン | タイプ |
|----|--------|-------|------|-------|-------|
| 東部 | 3032 | 1718 | DKK | 444 | 319 |
| | | | DKS | 2588 | 1399 |
| 西部 | 111815 | 39806 | HC | 5439 | 2942 |
| | | | PC | 98651 | 32208 |
| | | | SC | 7548 | 4528 |
| | | | VU | 177 | 128 |
| カシ | 508 | 433 | AD | 106 | 90 |
| | | | TS | 402 | 343 |
| 合計 | 115355 | 41957 | | | |

図表 4 各テキストの語数

² 語末の音と次の語の語頭の音がそれぞれの音の種類により変化して結合する現象。表記上一体化してしまい区分できない場合もある。

4. アパブランシャ語の地域差と格語尾

アパブランシャ語の中でも特に名詞の格機能を示す曲用語尾は多様な形式を有している。規範化を試みたヘーマチャンドラもその文法書『言語概説 (Śabdānuśāsana)』では様々な形式を並列している。たとえば名詞の中でもっとも数の多い a 語幹男性名詞は 8.4.331—339, 342, 346, 347 で規定され、図表 5 のようになる。

| | 単数 | 複数 |
|-----|--|------------------------------------|
| 主格 | deva, devā, devu, devo | deva, devā |
| 対格 | deva, devā, devu | deva, devā |
| 具格 | deveṇa, deveṇam, devem | devahim, devāhim, devehim |
| 属与格 | deva, devā, devasu, devāstu, devaho, devāho, devassu | deva, devā, devaham, devāham |
| 奪格 | devahe, devāhe, devahu, devāhu | devahum, devāhum |
| 所格 | devi, deve | devahim, devāhim |

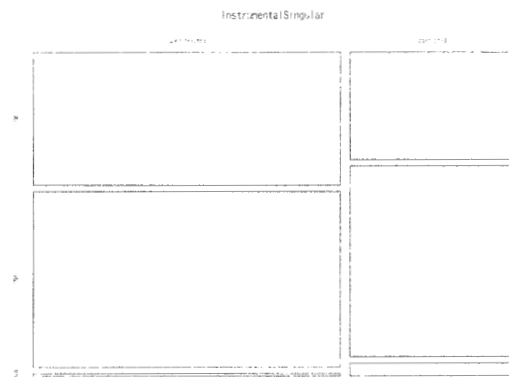
図表 5 a 語幹男性名詞の変化表

このように同じ機能に複数の語形が含まれることが多々見られ、同一作品内であっても複数出現することがある。そのため、言語特徴を分析する上で、基準となる形式が何であり、どういった条件で使い分けられているのかを突き止める必要がでてくる。

山畑(2011)では韻律の制限のために同一作品においても異なる形式が出現するのではないのかと考へ、調査した。コーパス中『パドマの行跡』を対象を絞り、分析を行った。韻律は特に各句の末尾に制限をかけることが多いため、格を示す語尾形式の中でも特徴的なものを用いて韻律との関係を探った。その数は図表 5 のようになる。属与格と奪格を一つにまとめているため、ヘーマチャンドラの解釈とは異なるが、校訂者の De Clercq の判断を優先した。図表 5 に見られる語尾形式のうち *-āṇa*, *-mmi* はごく少数のため無視することにしても、具格単数、与属奪格単数、与属奪格複数においては韻律の制限によって違いが出ているように思える。これらを図示すると、図表 7 から図表 9 のようになる。

| | | 制限無し | 制限有り |
|--------|------|------|------|
| 具格単数 | em | 1539 | 653 |
| | eṇa | 2011 | 1150 |
| | eṇam | 30 | 80 |
| 属与奪格単数 | ho | 2188 | 911 |
| | su | 254 | 170 |
| | ssa | 55 | 2 |
| 具格複数 | ehim | 1512 | 845 |
| | hu | 343 | 94 |
| | hum | 619 | 335 |
| 属与奪格複数 | ham | 278 | 158 |
| | āṇa | 9 | 20 |
| | mmi | 8 | 0 |
| 所格単数 | ahim | 332 | 54 |

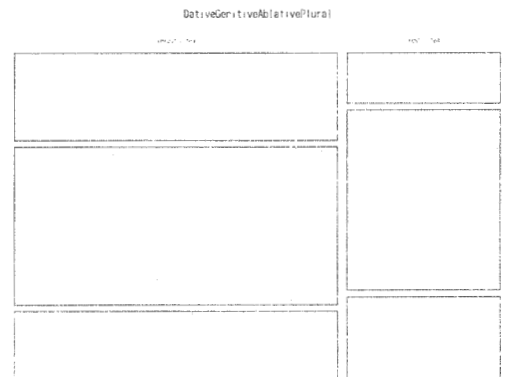
図表 6 韻律の制限と格語尾



図表 7 韻律制限の有無にもとづく具格単数の頻度



図表 8 韻律制限の有無にもとづく与属奪格単数の頻度



図表 9 韻律制限の有無にもとづく属与格複数の頻度

これらの図から韻律の制限がかかることによって次のような影響が出ていることが観察される。

- 具格単数 *-em* に対して *-ena*, *-enam* が優勢。
- 与属奪格単数 *-ho*, *-su* に対して *-ssa* が優勢。
- 与属奪格複数 *-hu* に対して *-hum*, *-ham* が優勢。

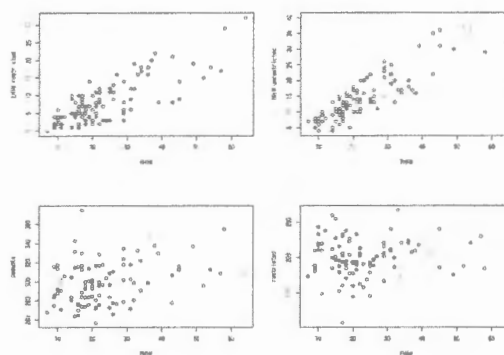
ただ、上記のような表現では全体数から判断されるのみで分布の様態までは表現できない。そこで山畑(2011)に加えてテキスト全体をいくつかのかたまりに分割して、それらのかたまりの性質がどのように分布するかを検討した。

『バドマの行跡』は複合語も一語と考えると総語数 98875 である。これを便宜上 100 個の語連続に分け、おおよそ 1000 語ごとのかたまりを作り、それぞれの中で語尾形式や韻律などの出現をカウントし、その分布との関係を調べた。

上記で併用される複数の格語尾が検討されなかった具格複数 *-ehim*、および所格 *-ahim* は比較対象がないため、図表 5 では出現数の多寡で検討しただけだが、テキスト全体を分割することにより、その分布状況がわかる。図表 10 と 11 はそれを図示したものである。横軸には一つのかたまりにおけるそれぞれの語尾形式の出現数、縦軸にはそれと関係があると思われる要素をとっている。4 枚の図のうち下の二枚はそれぞれ脚末の語数と韻律制限を受ける語数との関係を示している。*-ehim* も *-ahim* もその二要素とは関係が見られない。上二枚はそれぞれの形式のうち、韻律による制限の有無によって分け、それぞれの出現数をみたものである。すでに下の二枚から脚末という要素や制限という要素のみとは関係がないことがわかっているため、それぞれの語尾形式における制限の有無という観点から検討することができる。*-ehim* については制限の有無にかかわらず、正の相関をしめしている。それに対して *-ahim* は制限なしの場合には正の相関を示すが、制限ありの場合は相関を示しているとはいえない。

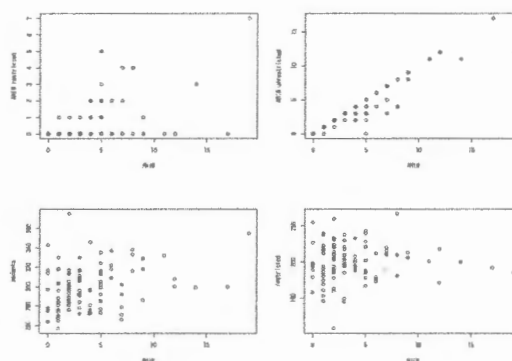
図表 10 と 11 から言えることは *-ehim* はどちらの場合でも同様に分布するが、*-ahim* は制限なしの場合のみ、全体の数に呼応しており、制限ありの場合にかなりランダムな出現をするということである。

このように韻律と語尾形式の間には一定の関係があることがわかる。そのため韻律情報を言語分析に利用することによってより詳細な情報を得られる可能性がある。



図表 10 具格複数 *-ehim* と他要素との関係

(左上: *-ehim* と制限付き *-ehim*、右上: *-ehim* と制限なし *-ehim*、
左下: *-ehim* と脚末語数、右下: *-ehim* と制限付き語数)



図表 11 所格 *-ahim* と他要素との関係 (それぞれの図の意味は図 4. *-ehim* と同様)

5. 学習による分類

現在コーパスに格納されているテキストについて現在存在するカテゴリー分類は未だ確定したものではない。そのため、何らかの形での検証が必要である。本稿では単純にテキストにおける単語の出現数に基づき、分類を試みる。手法として決定木およびナイーブベイズを利用した。本章では複合語を分割して利用する。

まず検証として『パドマの行跡』について構成の最小単位にあたるカダヴァカごとに文書を分割し、1414の文書を作成した。本作は全体で『ヴィドヤダラ・カーンダ』、『アヨーディヤ・カーンダ』、『スングラ・カーンダ』、『ユッタ・カーンダ』、『ウッタラ・カーンダ』の5部から構成されている。これについてまず決定木の手法により部別の学習を行った。200以上のカダヴァカに出現する語彙34語の出現数を因子として決定木を作成し、それによる自動分類を行った。

| | 予測 | | | | |
|------------|---------|---------|--------|------------|--------|
| 実際 | ayodhya | sundara | uttara | vidyadhara | yuddha |
| ayodhya | 342 | 0 | 0 | 0 | 0 |
| sundara | 0 | 210 | 0 | 0 | 0 |
| uttara | 0 | 7 | 48 | 16 | 154 |
| vidyadhara | 0 | 6 | 35 | 63 | 198 |
| yuddha | 0 | 2 | 31 | 17 | 283 |

図表 12 決定木による『パドマの行跡』部別の学習と予測結果

これを見ると『パドマの行跡』においては『アヨーディヤ・カーンダ』と『スングラ・カーンダ』は語彙から判別可能であるが冒頭の『ヴィドヤダラ・カーンダ』、および後半部の『ユッタ・カーンダ』と『ウッタラ・カーンダ』では語彙からは判別しきれないことがわかる。

さらに別の視点からの分類として『パドマの行跡』以外の7つのテキストをナイーブベイズの手法により学習させ、『パドマの行跡』の各カダヴァカがどの作品に分類されるかを見てみた。図表 12 がその結果である。結果からは時代や地域とは異なった分類が示唆される。『ハリセーナの行跡』と『ドーハーの宝庫(サラハ)』は地域的にはインドの西部と東部、宗教もジャイナ教と密教化した仏教というように異なる。また『ハリセーナの行跡』と『パドマの行跡』は同地域、同ジャンルで近いのは納得できるが、『サナクマラーの行跡』も同種のものであるにもかかわらず、分類されたカダヴァカ数はかなり少ない。

| タイトル | 地域 | 時代 | カダヴァカ分類数 |
|------|-------|----------|----------|
| HC | 西部 | 10世紀以降? | 507 |
| DKS | 東部 | 11~12世紀? | 495 |
| AD | カシミール | 9~10世紀? | 168 |
| DKK | 東部 | 7~12世紀? | 130 |
| VU | 西部 | 6世紀? | 106 |
| SC | 西部 | 12世紀 | 5 |
| TS | 西部 | 10~11世紀 | 4 |

図表 13 ナーブベイズによって作成した分類によって、『パドマの行跡』(9~10世紀)の各カダヴァカを配分した結果

この結果からは地域やジャンルよりもむしろ時代的な傾向が読み取れる。これはアパブランシャ語文献群内に存在する多様な差異が地域的な影響によるものではなく、文章語として整理されていた段階の違いによるものであるという考えに合致する。

このようにデータベースを利用することによって新たな視点からアパブランシャ語の分類を見直すことが可能となった。手法の適用方法をより洗練させることにより、さらに有益な結果がでてくることが予想される。

参考文献

- De Clercq, Eva. 2003. "Een Kritische Studie Van Svayambhūdeva's Paūmacariu."
- A Historical Syntax of Late Middle Indo-Aryan (Apabhraṃśa)*. 1998. John Benjamins Publishing Company.
- Tomoyuki Yamahata, 2012, The Classification of Apabhraṃśa —A Corpus-based Approach of the Study of Middle Indo-Aryan—, *Corpus-based Analysis and Diachronic Linguistics*, pp. 223-249.
- 山畑倫志, 2012, 「Suttanipātaの時代区分と韻律との関係」、『印度学仏教学研究』、日本印度学仏教学会、第60巻第2号、pp. 906-911
- 山畑倫志, 2011, 「韻律分析によるアパブランシャ語の格形式の確定」、平成23年3月、『印度学仏教学研究』(日本印度学仏教学会) 第59巻第2号、pp.832-839.