

小林正人・川村悠人・大山祐亮・張倩倩・塚越柚季

趣旨

これまで達成されてきたサンスクリット語の文法研究、とりわけヴェーダ文献（紀元前 1200 年頃から紀元前 500 年頃）で用いられるヴェーダ語の文法研究の成果は、印欧語比較言語学に負うところが大きい。この状況は今でも変わっておらず、印欧語比較言語学に基礎づけられたサンスクリット語研究はこれからも継続されていくだろう。一方、ドイツの Martin Haspelmath が 2005 年にオックスフォード大学出版局から刊行した『言語構造世界地図』（*World Atlas of Language Structures*, 2008 年にオンライン版が第 2 版として刊行）に象徴されるように、近年、言語類型論の進展が著しく、有益な研究成果が積みあげられている。そのような研究成果はサンスクリット語研究にも活用可能であり、新たな知見をもたらさうものである。さらに、自然言語処理など情報技術を活用した手法も人文学の領域で台頭してきている。サンスクリット語研究の分野でも、昨今の世界サンスクリット会議（World Sanskrit Conference）において情報技術を利用した言語研究に焦点をあてる専門部会が常に設けられているほどである。こうした技術の活用もまた、サンスクリット語研究に新たな地平を開きうる。

本ワークショップでは、上述した 1. 印欧語比較言語学、2. 言語類型論、3. 情報技術のそれぞれを活用した研究発表を通じて、これら三つの研究手法がサンスクリット語研究に対してどのように寄与するかを例示する。それによって、最新のサンスクリット語研究の現状を共有するとともに、未来におけるサンスクリット語研究としてどのような形がありうるかを探りたい。

そして、このような試みはサンスクリット語という個別言語の研究への貢献にとどまるものではない。第一の発表はこれまで法則性が見出されていなかった言語現象を説明する規則を定式化することで、第二の発表はサンスクリット語構文に言語類型論の一理論を当てはめてその通言語的妥当性を確立するための一事例を提供することで、第三の発表はコンピューターによる自然言語処理をサンスクリット文献に適用してその有効性を確証することで、全体として古典語研究のさらなる深化と広がり貢献するものである。

構成（計 2 時間）

企画者 川村悠人（広島大学） 司会者 小林正人（東京大学）

[1] 趣旨説明（5 分）**[2] 研究発表と質疑応答（発表 20 分 + 質疑応答 10 分、計 90 分）**

- (i) 大山祐亮（東京大学）「*dehi-edhi* の法則：印欧祖語*d^hに由来するサンスクリット *h*」
- (ii) 張倩倩（東京大学）「サンスクリットにおける受取手のとる格の選択基準—受益性と有生性に関する非対称性から」
- (iii) 塚越柚季（東京大学）「『リグ・ヴェーダ』讃歌の特徴を示す分散表現から得られる文書間類似度」

休憩（5 分）**[3] 全体討論（20 分）****各発表の要旨**

- (i) 「*dehi-edhi* の法則：印欧祖語*d^hに由来するサンスクリット *h*」

サンスクリット語においては印欧祖語の*d^hは原則としてそのまま dh [d^h]として保たれるが、散発的に h [h]にも変化することが知られている。これを h 化と呼ぶ。先行研究では、h 化は母音間の何らかの環境で起こるものと考えられていた。しかし、dehi「与えよ」と edhi「〜であれ」のようにほぼ同一の環境で h 化の有無が割れる例が存在するため、サンスクリット語の共時的な環境を前提にして h 化の条件を発見することは困難である。本発表では、サンスクリット語の前段階であるインド・イラン祖語の環境を前提とすると、h 化はインド・イラン祖語の喉音*H の直後で起こると定義できる (dehi-edhi の法則)、と主張する。これによって、dehi (<*da-dH-d^hi) と edhi (<*(H)az-d^hi) のような例の h 化の有無が予測可能となる。なお、印欧祖語の子音間の*H のサンスクリット語における反映形が、最も無標な母音*a ではなく前舌母音 i であることから、これは順行口蓋化の一種であると思われる。

(ii) 「サンスクリットにおける受取手のとる格の選択基準—受益性と有生性に関する非対称性から」

サンスクリット語には、最古の文献『リグ・ヴェーダ』(紀元前 1200 年頃) から śās「教える」等の二重対格をとる動詞が存在し、「言う」や「運ぶ」などの動詞も時代が下るにつれて二重対格をとるようになった。これらの動詞はいずれも通言語的に二重対格をとりやすいとされるものである (Martin Haspelmath, “Ditransitive Constructions,” *Annual Review of Linguistics* 1: 19–41, 2015, p. 37)。一方で、通言語的に最も二重対格をとりやすいとされる (ibid.) 動詞 dā「与える」は、一貫して受取手が与格で、移動体が対格で表される。本発表は、受取手と移動体の有生性に関する非対称性 (例えば二者がともに人間なら非対称性は低い) と受取手の受益性という指標を用いて、この状況が生じる理由を説明する。調査の結果、『リグ・ヴェーダ』では受取手と移動体の非対称性が高ければ受取手が対格で、低ければ受取手が与格で表される傾向があること、紀元前 5 世紀頃からは受取手の受益性が低い場合に受取手が対格で表されるようになって二重対格が拡大したことが判明した。『リグ・ヴェーダ』における二重対格の分布には有生性に関する非対称性の高低が、そして古典期に動詞が二重対格をとるようになるかどうかには受益性の高低が、それぞれ基準となっていると言える。「与える」が一貫して二重対格をとらないのは、受取手と移動体の非対称性が低く、かつ受取手の受益性が高いからであると結論づけられる。

(iii) 「『リグ・ヴェーダ』讃歌の特徴を示す分散表現から得られる文書間類似度」

サンスクリット語文献『リグ・ヴェーダ』の各巻は、複数の詩節から成る讃歌の集合である。讃歌や詩節ごとに、創作者である詩人や神格が異なる。

Doc2Vec とは、文書の ID と単語列をもとにその列に続く単語を推測するアルゴリズム (Paragraph Vector: A distributed memory model, PV-DM) と、文書の ID をもとに文書内の単語列を推測するアルゴリズム

(Paragraph Vector: Distributed bag of words, PV-DBoW) とを統合したものである。単語の意味を表現するベクトルは、共起する前後の数単語とその頻度から学習される。現実の言語使用のデータをもとにして一種の言語知識を得るという点で用法基盤モデルの考え方に近い。単語ベクトルが結合された文書ベクトルにより、文書間の類似度の計算や、新規文書が既存文書群のいずれの類に属するかの推測などができる。

Doc2Vec を用いて計算した『リグ・ヴェーダ』の各讃歌や詩節間の類似度を、グラフによって視覚化すると、類似度の高い讃歌や詩節群あるいは低い群を明確に示せる。類似度グラフから、例えば単一の神格への讃歌から成る『リグ・ヴェーダ』第 9 巻が他より類似度の高い群であることを確認できるように、Doc2Vec が『リグ・ヴェーダ』中の内容や言語特徴などが異質な讃歌を確かに導出できることを本発表において示す。