

平成 30 年 5 月 15 日現在

機関番号：14401

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K02517

研究課題名(和文) ウェブコーパスを利用したヒンディー語・日本語の複合動詞の対照研究

研究課題名(英文) A comparative study of compound verbs in Hindi and Japanese using a large web corpus

研究代表者

西岡 美樹 (NISHIOKA, Miki)

大阪大学・言語文化研究科・准教授

研究者番号：30452478

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：本研究は、現代ヒンディー語の大規模ウェブコーパスを構築し、コーパスから得られる使用頻度等の統計的指標を元に、ヒンディー語の複合動詞について、語用論の面から日本語との対照研究を行うものである。

主な成果はこれまで定説であった「複合動詞のV2('jaanaa' GOや'dena' GIVE等)と否定辞が基本的に共起しない」について、実際はこれらV2が否定辞と共起している事実を量的に例証したことである。さらにこの否定辞と共起しているV2の例について、日本語の補助動詞「しまう」と否定辞が共起する環境を援用しつつ、ヒンディー語のV2の共起する環境が「しまう」のそれに酷似していることを明らかにした。

研究成果の概要(英文)：The aim of this research project is to verify, by using a large web corpus, Jagannaathan's established claim that secondary verbs (V2s) in compound verbs (CV: V1+V2) very rarely co-occur with negative markers in Hindi. The most important results of the research project are as follows:

V2s such as Hindi verbs jaanaa GO, denaa GIVE, and lenaa TAKE, together with negative markers, tend as a whole to behave quite similarly to Japanese verb shimau PUT AWAY - one of most frequent Japanese auxiliary verbs, which renders an aspect of 'completeness'. The examples from the corpus indicate that the possible environments for co-occurrence with the negative markers are essentially limited to clauses that carry some additional semantic element, such as an imperative (periphrastic expression), subjunctive, conditional, etc. Co-occurrence with the indicative is rare.

研究分野：言語学、ヒンディー語学、コーパス言語学

キーワード：複合動詞 ヒンディー語 コーパス 日本語 completeness benefactive 補助動詞

1. 研究開始当初の背景

(1) 言語研究において、考察対象となる言語資料の主な抽出法には、①研究者自身の内省(母語の場合)または任意の母語話者の内省(外国語の場合)による判断、②アンケートによる適正調査、③コーパスによる大量言語調査、がある。言語研究は①の手法で主に行われてきた。しかし①の場合、母語話者が必ずしも内化された言語知識を完璧に意識化できるわけではないから、結果が信頼性に欠ける。②の場合は①よりも母集団が増えるが、③の大量言語調査から得られる結果と比べ、母集団の偏りは否めず、信頼性が高いとはいえない。これらはそれぞれ一長一短なところがあるためいずれかが万全な方法なわけではなく、あくまで相補的に用いられるのが言語研究の精度を高めるためには理想的だが、世界で屈指の話者数を抱えるヒンディー語は、これまで③がほとんど整備されてこなかった。

(2) ヒンディー語は 19 世紀後半から現在に至り、非ヒンディー語母語話者(以下、非母語話者)である外国人研究者による研究や母語話者による研究が続いているが、前者は可視化された形態、統語面での分析にとどまっておき、非母語話者による文脈との関係を重視した語用論的研究はあまりない。その理由として、まずこれらの分野を研究するには当該言語の高度な言語知識と言語運用能力並びに母語話者相当の言語経験を要することが挙げられる。一方、母語話者は既に母語の運用能力を備え自由に会話できるため、その用法上の仕組みを意識化させにくい。この点、非母語話者は、仮に長年の言語経験があったとしても、母語話者未満であるため任意の言語現象に関して直観や内省による例証が困難である。

(3) かつて研究代表者(以下、代表者)が従事した動詞連結に関する科研課題で、これまでの代表者自身の言語経験から「本動詞(V1)に何らかのニュアンスを付加するのに概ね相当し、話し言葉(会話体)に比較的出現しやすい」という仮説を立て、ヒンディー語の会話体を含む短編小説数編について、日本語訳を調査し、ヒンディー語の複合動詞(V1+V2)のうち、語幹接続のV1に続くV2の *jaanaa*「行く」、*denaa*「与える」、*lena*「取る」が日本語のテ形接続の補助動詞「しまう」、「いく」、「やる(あげる)」、「もらう」等に訳される傾向にあることを指摘した。これは手動による原文 vs. 対訳のパラレルコーパス手法だが、量的に限られたものになり母語話者、非母語話者の研究者いずれに対しても、十分に例証するには至れなかった。他方、元々ヒンディー語の非母語話者である Jagannaathan の *Prayog aur prayog* (1981) が「ヒンディー語の複合動詞は否定辞と共起しない」という用法上の制限について指摘していることを敷衍し、日本語の補助動詞で使用頻度が高く、日本語の対訳にも出現しやす

い「しまう」について、国立国語研究所のコーパス BCCWJ を利用し、試験的に否定辞と共起する頻度と共起する例を詳細に調べ、実際どのような共起制限がかかっているか、共起を許す場合の環境を詳細に分析した。対照研究を行うにはこの結果と対比し、ヒンディー語の複合動詞群の使用頻度、否定辞との共起等に関するデータを出す必要があったが、先述の通り、ヒンディー語の場合、言語研究に適した検索環境がほとんど整っていないという状態であった。

2. 研究の目的

本研究は、言語研究に適した現代ヒンディー語のウェブコーパス(既存の Sketch Engine を超える数億語規模)を構築し、非ヒンディー語母語話者の研究者による語用論的研究の足場を整え、ヒンディー語をはじめ南アジア諸語で盛んに研究された複合動詞について、語用論の面から日本語との対照研究を行うことを目的としたものである。コーパスから得られる頻度等の統計的指標によりこれまで非母語話者である研究者では容易に扱えなかった語用論的研究、特に南アジア諸語でしばしば取り上げられる文脈依存率の高い複合動詞についてジャンルによる出現率の差や実際の用法(運用面)等を明らかにすること、これにより将来的に世界レベルでヒンディー語研究及び南アジア諸語の研究分野の裾野を広げることに寄与し、さらにそれらの分野の研究が外国語としてのヒンディー語教育にも貢献することを目指したものである。

3. 研究の方法

本研究は、ウェブコーパス構築と複合動詞の研究を同時進行で行う。

(1) コーパス構築

初年度にウェブコーパスにアノテーションを行い、コーパス構築並びに簡易版の検索スクリプトの開発を行う。次年度には、国内外の研究協力者から得られたフィードバックを元に、コーパスのアノテーション等の修正を行い、さらに、公開版の検索システムの開発も行う。これらは、代表者の研究で得られた検索結果をいつでも誰でも再現し、検証できるように、本研究の研究期間終了までに、国内外に無料で一般公開する。

(2) 複合動詞の研究

先述の通り、初年度の前半に言語研究のためのパイロット的検索環境を整え、複合動詞の研究に取り掛かる。具体的にはウェブコーパスを活かし、ヒンディー語の主要な複合動詞について検索をし、複合動詞のうち頻度の高いものをいくつか選定し、本格的に調査、分析する。また、日本語の複合動詞(正確にはテ形接続の「しまう」や授受動詞のような補助動詞群)との対照研究も行う。

4. 研究成果

(1) コーパス構築

初年度は、ヒンディー語のウェブコーパス構築にかかる準備を技術面の研究協力者である赤瀬川 (Lago 言語研究所代表) と行った。具体的には本ウェブコーパスの規模、収集法の詳細、POS タガーの選定、手順の確認を行った。次に、ウェブデータの収集に際して必要な作業 (シードの作成とヒンディー語以外のウェブデータと重複データの排除) を行い、ウェブデータの質を確認した。また、後に開発する本コーパス用の検索ツールのプロトタイプである簡易検索スクリプトの作成を行い、同年 12 月に試験運転、コーパスと検索スクリプトの動作確認をした。その検索結果を元に、次年度の開発に向けての打ち合わせや検索に付随する問題について協議した。本プロジェクトの課題であるヒンディー語の複合動詞について検索を行ったが、その際に品詞タガーでタグ付けされたレマに新たな不備が見られた。この点は次年度の改良点とした。

次年度は、ヒンディー語のウェブコーパス (Corpus Of Spoken Hindi: COSH) の検索テストで判明したアノテーションの問題について打ち合わせを行い、解決法を議論した。開発面での柱であった検索インターフェース (COSH Concordancer: COSH Conc) については、同年 10 月公開を目指した開発計画と、ウェブコーパス公開の準備 (ドメイン取得、レンタルサーバー契約など) に関する詳細な打ち合わせを行った。7 月にはウェブサイトの公開、COSH のトップページ、実際の検索に必要なユーザーマニュアルの作成とその英語化についての打ち合わせも行った。その際、COSH & COSH Conc を載せるレンタルサーバーについて、ヒンディー語で使用するデータベースの容量がラテン文字よりも容量が大きいと、当初見込みよりも容量が大きいサーバーをレンタルしなければならないことが判明した。9 月には Preliminary Version を内部公開し動作確認を行い、海外の研究協力者たちからのフィードバックを得て、11 月には正式版 (Version 1.00) を海外の南アジア諸言語の研究者用メーリングリスト上で一般公開 (日本国内向けには日本南アジア学会のメーリングリストで 12 月に一般公開の広報済) した。

最終年度は、前年度に開発したヒンディー語のウェブコーパスに関わる問題 (アノテーション、コーパスのデータやウェブサイトに関する修正点等) について、技術面の研究協力者や海外の研究協力者らと対面あるいはメール等で打ち合わせを行い、将来的な修正点を検討した。また、ウェブコーパスの現地 (ヒンディー語母語話者) でのさらなる普及を目指し、研究協力者である赤瀬川と、ヒンディー語のインフォーマントとしても代表者の研究に協力している Srivastava

(Central Institute of Hindi, retired professor) と Ms. Sharma (Japan Foundation Delhi 他、フリーランス日本語教師) とともにウェブサイトのヒンディー語ページを作成し、2018 年 2 月上旬に一般公開した。

なお、COSH & COSH Conc は一般公開から既に約 1.5 年経過するが、研究開始時の予想に反して、言語研究者である我々のような非ヒンディー語母語話者 (つまり日本も含むインド国外) からのアクセス数よりも、インド本国からのアクセス数が多い。また、次にアクセスが多いのはインド系移民や在外インド人 (NRI) も多い米国であることは、注目に値する。

(2) 複合動詞の研究

初年度に、まず研究開始前年度に、ウェブ公開されている新聞記事を元に研究協力者の赤瀬川が実験的に作成したコーパス (約 1 千万語) を利用し、日本語のテ形+「しまう」と否定辞の共起制限の結果と対照させた “Restrictions on co-occurrence of ‘STEM + *jaanaa*’ and negation in Hindi: a contrastive analysis with ‘-*te* + *shimau*’ in Japanese” を SALA-31 (Lancaster University, UK) で発表した。これは前出の Jagannaathan (1981) の「否定辞と複合動詞は基本的に共起しない」「命令表現の場合例外的に共起する場合がある」等の主張を取り上げ、ヒンディー語で最も使用頻度が高い複合動詞 *jaanaa* (原義は「行く」) が本当に否定辞と共起しないのか、もし共起するとしたらどのような環境で共起するのかを、コーパスを使って客観的に調査し分析したものである。主な結果は、①ヒンディー語の *jaanaa* と否定辞の共起頻度は低い、②直説法の文では否定辞との共起はまず起こらない (つまり言い切りの文を否定する時には共起しない)、③命令表現では否定辞と共起することがある、④副詞節、関係節等の節内ならば、十分に否定辞と共起される、の 4 点であった。これらの結果は日本語のテ形+「しまう」の否定形の場合とほぼ同じであった。

ウェブコーパスと検索スクリプトができあがった段階で、*jaanaa* について、さらに大規模なウェブコーパスで検証実験を行った。初年度の研究で得られた知見と合わせ、次年度の 4 月に “Functions of *jaanaa* as a V2 in Hindi: From Lexicalization to Grammaticalization” を SALA-32 (University of Lisbon, Portugal) で発表した。今回のウェブコーパス上での調査では、直説法の文と考えられるもので、*jaanaa* と否定辞が共起している例が一定数見つかったのである。この事実は、これまでのヒンディー・ウルドゥー語のみならず南アジア諸言語の研究での定説を揺るがすものになった。見つかった例文すべてについて文脈も合せて検証することは不可能なため、無作為抽出したものをいくつか検証したが、否定辞と共起している第

一動詞 (V1) に偏りはあるものの、共起する例が見つかっている。代表者はこれらの抽出例から *jaanaa* には文法化しているものと、V1 と合せて語彙化しているものがある可能性があることを本発表で指摘し、さらに否定辞が影響する範囲 (スコープ) の問題も関わっていることにも言及した。これについては、今後母語話者の協力を得ながら検証例を増やし、研究を深める必要がある。

一方で、チベット・ビルマ諸語の研究者である Dr. Lahaussais (Histoire des Théories Linguistiques, CNRS) の協力により、5月に “A Corpus-based comparative study of V1+V2 concatenations in Hindi and Japanese: Co-occurrence of STEM + *jaanaa* / -te + *shimau* and Negation” を Université Paris Diderot で講演した。これは、日本語やヒンディー語の複合動詞のような統語から語用論にまたがる研究には、大規模コーパスが有用であることを主張したものである。意見交換の際に、研究に必要なコーパスの語数の問題やコーパスの質がどの程度保証されるのか等、その有用性に疑問を唱える意見も出たが、それはこれからの課題であり、いずれにしても母集団が寡少な対面調査や文献調査では、研究に限界が出てくること、その限界を補えるものの一つが規模の大きいコーパスであることを代表者は主張した。

次年度後半はこれまで扱った *jaanaa* の研究を一旦おき、ここまで得られた新たな知見を踏まえながら、*jaanaa* 同様、COSH & COSH Conc 上 (この時点では未公開版) を使い、*denaa* (原義は「与える」) の使用頻度と共起する環境を調査した。この調査結果を、最終年度の5月に “Does *denaa* GIVE as a V2 render benefactive meaning in Hindi?: A corpus-based comparative analysis in Hindi and Japanese” と題して、SALA-33 (Adam Mickiewicz University, Poland) で発表した。日本語では補助動詞の「やる／あげる」が否定辞と共起しても問題なく、その場合、受益 (benefactive) の意味を否定することになるのは、日本語母語話者なら直観的に理解できる。しかし、対するヒンディー語について COSH & COSH Conc 上で調査すると、先の *jaanaa* の②「直説法の文では否定辞との共起は起こらない」が基本である。しかし、*jaanaa* 同様、低い頻度ではあるものの例外が見られた。それらのうち無作為抽出をしたものを観察すると、否定辞とは別に小詞 (日本語のとりたて詞あるいは副助詞／係助詞の類) を伴っていることが判明した。つまり、*jaanaa* 同様、否定辞が及ぼす否定の範囲にも関わりがあることを示す結果となった。一方で、ロシア人の Liperovskii (1984) の *Hindi language -- Verb* でも複合動詞と否定辞との共起制限について言及があるという情報を得たため、Prof. Pande (Jawaharlal Nehru University, retired professor) の協力でもロシア語から英語に翻訳、今後の資料として

参考にすることにした。

さらにこの続編として、同年7月には、代表者の教え子でもある、シンハラ語と日本語を共に母語とする尾濱 (大阪大学学部生) の協力を得、いわゆるインド・アーリア諸語でヒンディー語の姉妹語とされるシンハラ語の「与える」相当の補助動詞の振る舞いと比較した “Does *denaa* GIVE as a V2 render benefactive meaning in Hindi?: A comparative case study of Hindi, Japanese and Sinhalese” を、言語科学会第19回国際年次大会 (京都女子大学) で発表した。結果、シンハラ語は日本語と同じで、補助動詞の「与える」自体を否定でき、その場合の意味も、日本語と同じく受益の意味を否定していることが判明した。対するヒンディー語は、*denaa* を伴う複合動詞のまま否定することができない。つまり、日本語的な「して・あげない」相当がいえないのである。否定文を作る場合は、V2 の *denaa* を取り除いて V1 のみを否定しなければならない。この事実は、ヒンディー語の *denaa* は、あくまで *jaanaa* に似た完遂 (completeness) の意味合いを示すものであり、受益 (benefactive) という意味合いが含まれていない可能性が高いことを示唆している。また、裏を返せば、例えばヒンディー語からこの複合動詞を日本語に訳す場合、この「受益」の意味を日本語母語話者が読み込み過ぎている可能性があることも指摘した。さらに、この点について、ヒンディー語と同族系の言語とされるシンハラ語が、ヒンディー語でなく日本語に酷似していることも同時に明らかになった。

最終年度後半には、前年度ウェブコーパスから *denaa* と併せて収集した *lenaa* 「取る」と *rakhnaa* 「置く」のうち、*lenaa* 調査を行っている。結果は本研究課題終了後の平成30年度に行う予定である。

なお、このウェブコーパス開発とヒンディー語の言語研究への応用について、研究協力者の Dr. Désoulières (INALCO/CERLOM/ASIEs) の協力の下、研究協力者の赤瀬川と共同で、平成28年12月に “The Development of a Web Corpus of Hindi Language and Corpus-based Comparative Studies to Japanese” と題して WSSANLP-6 (GranCube Osaka, Japan) で発表した。また、本研究の知見は、『現代ヒンディー語文法概説 初級～初中級編』の執筆の際、「複合動詞」の項目でも反映させている。

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1件)

- ① Miki Nishioka, Shiro Akasegawa, The Development of a Web Corpus of Hindi Language and Corpus-based Comparative Studies to Japanese, 6th Workshop of

South and Southeast Asian Natural Language Processing Proceedings of the Conference, 査読有、2016、114-123
<http://www.aclweb.org/anthology/W16-3712>

[学会発表] (計 5 件)

- ① Miki Nishioka, “Does *denaa* GIVE as a V2 render benefactive meaning in Hindi?: A comparative case study of Hindi, Japanese and Sinhalese”, 言語科学会第 19 回国際年次大会 (JSLS 2017)、July 2, 2017
<http://jsls.jp/con/2017/>
- ② Miki Nishioka, “Does *denaa* GIVE as a V2 render benefactive meaning in Hindi?: A corpus-based comparative analysis in Hindi and Japanese”, 33rd South Asian Languages Analysis Roundtable (SALA-33)、May 16, 2017
http://sala33.home.amu.edu.pl/?page_id=35
- ③ Miki Nishioka, Shiro Akasegawa, “The Development of a Web Corpus of Hindi Language and Corpus-based Comparative Studies to Japanese”, 6th Workshop of South and Southeast Asian Natural Language Processing (WSSANLP-6) in the 26th International Conference on Computational Linguistics (COLING)、December 11, 2016
<http://www.sanlp.org/wssanlp2016/>
- ④ Miki Nishioka, “Functions of *jaanaa* as a V2 in Hindi: From Lexicalization to Grammaticalization”, 32nd South Asian Languages Analysis Roundtable (SALA-32)、April 28, 2016
<http://sala-32.wixsite.com/home/programme>
- ⑤ Miki Nishioka, Shiro Akasegawa, “Restrictions on co-occurrence of ‘STEM + *jaanaa*’ and negation in Hindi: a contrastive analysis with ‘-te + *shimau*’ in Japanese”, 31st South Asian Languages Analysis Roundtable (SALA-31)、May 16, 2015
<http://ucrel.lancs.ac.uk/sala-31/programme.php>

[図書] (計 1 件)

- ① 西岡 美樹, 東京外国語大学アジア・アフリカ言語文化研究所、『現代ヒンディー語文法概説 初級～初中級編』、2017、77
<https://publication.aa-ken.jp/>

[産業財産権]

○出願状況 (計 件)

名称 :
発明者 :
権利者 :
種類 :
番号 :
出願年月日 :
国内外の別 :

○取得状況 (計 件)

名称 :
発明者 :
権利者 :
種類 :
番号 :
取得年月日 :
国内外の別 :

[その他]

ホームページ等

COSH and COSH Conc [English version]

<http://www.cosh.site/>

COSH and COSH Conc [Hindi version]

<http://www.cosh.site/hi/>

6. 研究組織

(1) 研究代表者

西岡 美樹 (NISHIOKA, Miki)

大阪大学・大学院言語文化研究科・准教授

研究者番号 : 30452478

(2) 研究分担者

()

研究者番号 :

(3) 連携研究者

()

研究者番号 :

(4) 研究協力者

赤瀬川 史朗 (AKASEGAWA, Shiro)

尾濱 智子 (OHAMA, Chiko)

Gunjan Sharma

Ashwani Kumar Srivastava

Hem Chandra Pande

Aimée Lahaussais

Alain Désoulières