

平成 30 年 6 月 4 日現在

機関番号：13901  
 研究種目：基盤研究(C) (一般)  
 研究期間：2014～2017  
 課題番号：26370483  
 研究課題名(和文)大規模コーパスに基づく名詞と形容詞の使用パターンと構造化に関する日仏語対照研究

研究課題名(英文)The Usage Patterns of Nouns and Adjectives and their Constructionalizations: A Contrastive Study of Japanese and French based on Large-scale Corpora

研究代表者  
 藤村 逸子(Fujimura, Itsuko)  
 名古屋大学・人文学研究科・教授

研究者番号：50229035  
 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)：日仏語の大規模コーパスをもとに100万件規模のNグラムデータベースを構築した。このデータを対象に、種々のコロケーション指標を評価し、頻度結合の強度(Log-r)により2グラムの特徴付けが可能であることを示した。同時に、データの観察により、以下の言語事象を発見し、コーパスに基づき詳細に記述・説明した。1)フランス語の「名詞+名詞」と「名詞+de+名詞」間の変化と変異、2)フランス語の「femme+人間名詞」の語順は、後続名詞の女性名詞化のためのfemmeの接頭辞化に由来、3)フランス語の親族や王・王妃などを表す名詞の形容詞化において性別が文法上の性として機能する現象。

研究成果の概要(英文)：We built the database of 1 million Ngram from the very large corpora of French and Japanese. After evaluating several collocation indices with these data, we have demonstrated the measurement of the characteristic of bigrams according to their frequency and their degree of compositionality measured by their log-r. At the same time, through the observation of the data, we discovered some linguistic facts very little or never treated in the literature and described and explained them in detail: 1) changes and variations between <NN> and <N de N> in French, 2) typologically irregular word order <femme + human noun> can be explained by the fact that femme is a prefixoid feminizing the sequence 3) transfer of 'sex' into grammatical gender in the adjectivation of kinship and loyalty terms in French.

研究分野：言語学、フランス語学

キーワード：言語変化 構文化 コーパス コロケーション 文法上の性 統計手法 日本語 フランス語

### 1. 研究開始当初の背景

従来の言語研究はパラダイムへの関心が主流であり、語彙項目の文法パターン(コリゲーション)やコロケーションパターンなど、共起に関する問題は注目されてこなかった。その理由の一つは、意味や文法性の差は内省による分析が可能とされるのに対し、文法パターンやコロケーションパターンは無意識に属し、無限に存在する語彙項目の一つ一つの使用パターンを発見して一般化することには、方法論上の困難があったからである。しかし、コンピュータの技術革新に伴い、大規模データを用いて、語彙項目の一つ一つに対して、詳細な使用のパターンを観察することが可能になり、この分野の研究が次々と発表されている。しかし一方で、言語研究がコンピュータの処理能力に追いついていないのも確かである。同じ一人の研究者、あるいは一つの研究チームにおいて程度の差はあっても、言語学と情報科学の両方の分野を理解する研究者が協力して、従来の方法とコンピュータを用いた方法を融合させた研究が望まれる。

### 2. 研究の目的

本研究は、言語現象を人間の動的な言語使用の痕跡と見なす立場に立ち、書き言葉と話し言葉のコーパスを用いて、日本語とフランス語の名詞句内のコロケーションとコリゲーションのパターンを定性的および定量的に観察し、言語使用と言語構造との関係の解明を目指す。書き言葉コーパスからは、種々のジャンル・時代における言語使用の大規模なデータを採取できる。話し言葉コーパスからは発話者の発話意図により直結した、言語使用の精密なデータが得られる。この両者を組み合わせることによって、言語使用の傾向とその構造化の過程を、使用頻度、意味領域、品詞などの多因子を仮定して推定する。また日仏語の対照から、汎言語的な一般化を目指す。

### 3. 研究の方法

日仏両語の大規模書き言葉コーパスを拡充し、Nグラムなどの大規模データを作成する。日仏両語に共通した意味領域(人間、サイズ、色など)を選定し、名詞+名詞、名詞+形容詞、名詞+助詞+名詞などの名詞句を選定し、頻度の高い表現をコーパスから抽出する。コロケーションパターン、コリゲーションパターン、意味パターン(語と意味間)の強度と頻度を計算する。その他の情報を加えてデータベースを作る。データベースの定量的・定性的な観察により日仏語の特徴を明確にする。現象を確率論的にとらえる根拠を理論化し、言語使用が言語構造を作り上げるモデルを作成する。すでに構築済みの話し言葉コーパス『名大会話コーパス』の利用方法を改善して、話し言葉をデータとして有効に利用するよう

にする。

### 4. 研究成果

研究の主要な成果は以下の3点である。

1. 日仏語の大規模コーパスをもとにして、多量の品詞タグ付きのNグラムデータベースを構築した。共同研究によりすでに構築済みの英語Nグラムと併せて、本研究課題において利用した。また、これらのデータは今後の研究においても利用が可能である。
2. 日本語・フランス語・英語のそれぞれ100万件規模の2グラムデータをもとにした、統計手法に基づく2グラムの特徴付け。コロケーションの指標として多用されてきた、MIスコア、tスコア、Dice, Jaccardなどを本研究で提案したLog-rと比較して、その欠点を明らかにした。2語連語は、結びつきの強度と頻度によって特徴付けられる。MIスコア、tスコア、Dice, Jaccardはいずれも、結びつきの強度のみをはかる指標ではないために、強度と頻度の2次元的な特徴付けには向かない。一方、Log-rは強度のみを測る簡素な指標である。結びつきの強度が強く頻度が低い結合は意味的に1語化するが、頻度の高い結合は結びつきの強度は必ずしも強くはなく、文法的・機能的な連語に相当する。いずれにせよ、全ては連続体を構成しており、1語なのか2語なのかなどに関して境界が定まっているわけではない。
3. Nグラムデータベースの観察により、過去に記述がないか、あるいは十分には記述されていないが、言語学的に重要な研究課題を発見することができた。それぞれの課題について、各言語のコーパスを用いて、文脈の観察を入念に行い、その使用を詳細に記述した。言語の使用に基づき、構造化の過程を推測することができた。以下に、主な5つの研究課題に関して詳述する。

(ア) フランス語の「名詞+名詞」と「名詞+de+名詞」の使い分け。フランス語では「名詞+名詞」の連続は、英語の影響などで頻繁に起こると言われているが、前置詞の消去は、一般に述べられているほど頻繁ではなく、同じ名詞が「名詞+名詞」と「名詞+de+名詞」の両方の形で出現するのはきわめてわずかである。ほとんどの場合、スキーマは前もって定まっていると言える。*Effet domino vs Effet de serre* の対立は意味的に説明が可能である。Fin janvier vs fin de janvier>など、構造変化が進行中の

ために、両方の使用が可能な連続もあるが、必ずしも、「名詞+名詞」の方向に向かっているとは言えない。

(イ) フランス語の「femme + 職業名詞」の語順は日本語と同じであり、言語類型論的観点からはイレギュラーである。Femme の前置は、職業名詞を女性名詞にする一般の女性の接尾辞と同様の役割を果たし、半ば接辞化して使われていると考えられる。フランス語の「名詞+名詞」の連続では前置名詞の文法上の性が連語の文法上の性を決定することにより、femme を前置する慣用が生じたと考えられる。日本語は、「女性医師」「白人女性」のように両方の語順があり入れ替えは不可能であるが、日本語の語順は意味論的に説明が可能である。外延が広く内包の大きい概念は形容詞的に使われやすく、逆の場合には名詞的に使われやすい。「医師」と「白人」を比べた場合、医師は白人よりも内包が大きく、外延は小さい。

(ウ) フランス語では、親族名詞や王・王妃を表す名詞が比喩的に前置名詞を修飾する際に、ville sœur, pays frère のように、名詞の表す性別が形容詞の文法的性と同様の振る舞いをして、共起する名詞を製薬する。男を示す語(frère)は男性名詞を、女を示す語(soeur)は女性名詞とのみ共起し、pays とほぼ同義の nation は nation soeur のみが可能で、nation frère は不可能である。これまでに記述のなされていないこの現象を、vie étudiante, mouvement étudiant のようなよく知られた人間を表す名詞の形容詞化と対比してデータに基づき記述・説明した。étudiant/étudiante は一般に、名詞用法以外に、形容詞用法があると認定されているが、人間名詞はすべて形容詞としても用いられるわけではなく、さらなる研究が必要である。日本語では、「親会社」「子会社」「孫会社」「親指」など、共起名詞の制約は意味論的に透明性が高いが、フランス語では、名詞の性と人間の性の関連の故に不透明性であると言える。

(エ) コーパスに基づき、日本語の「甘い」の意味記述を行い、フランス語や他の多くの言語とは異なり、この語が女性性と密接な関連があ

ること、また、コーパスに基づき、日本では「甘さ」は男らしさの欠如と結びつきが深いことを示した。男らしさは、「辛さ」と共起しやすく、従って、「甘さ」は日本語においては、否定的なコノテーションを持っている。

(オ) コリゲーションパターンの研究として、フランス語の使役構文における、被動作主を表す à と par の対立に関する研究をコーパスに基づいて行った。

話し言葉を専門とする研究者と共同して、話し言葉コーパスの「名大会話コーパス」の分析を音声データも利用して行い、イタリアの学会で紹介した。話し言葉における使用頻度は、構造化の観点から近年重要視されているために、研究のブレイクスルーとして期待できる。また、「名大会話コーパス」を国立国語研究所のサイトに公開し、国立国語研究所が開発する検索ツールが使えるようになった。このようにして、今後の研究に役立てるための準備を行った。

## 5. 主な発表論文等

〔雑誌論文〕(計 5 件)

1. Fujimura, Itsuko, 2018 印刷中, «L'énigme de l'ordre des mots : «femme + noms d'humains», Linx, revue des linguistes de l'Université Paris Ouest Nanterre La Défense, 査読有
2. 藤村逸子, 青木繁伸, 2017, 結合の強度を測る指標としての Log-r の有用性 : 日・英語のバイグラムデータに基づく MI, LLR などとの比較, 『言語資源活用ワークショップ発表論文集, 1 巻』, pp.365-376 査読有 doi.org/10.15084/00001492
3. 藤村逸子, 2016, フランス語の使役構文における被使役者の表示 : a vs par - 実例に基づく検証, 『フランス語学研究 50 別冊号』(頁 : 119-138) 査読有
4. Fujimura, Itsuko, Aoki, Shigenobu, 2016, A New Score to Characterise collocations: Log-r in Comparaison to Mutual Information in Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives, Corpas Pastor, G. (Ed.), Tradulex. http://www.euophras2015.eu/euophras2015\_bookoffullpapers/! 査読有
5. 藤村逸子, 2015, 「フランス語質問箱 : 文法上の性の話」『フランス語学研究』49 号, pp. 145-149, 査読有

〔学会発表〕(計 8 件)

1. Fujimura, Itsuko 2017, Similitude apparente mais structure différente en français et en japonais, *ville sœur, femme gouverneur*, 『文法化、語彙化、凝結：日仏対照言語学的アプローチ』、名古屋
2. Emanuela Cresti, Fujimura Itsuko, 2017, The information structure of spontaneous spoken Japanese and Italian in comparison: a pilot study, 2017: Li Congresso Internazionale SLI. ナポリ (イタリア)
3. Fujimura Itsuko, 2017, *Villes soeurs et pays frères: le sexe des substantifs transféré en genre grammatical dans leur adjectivation*, Colloque international; L'adjectivité, パリ (フランス)
4. 藤村逸子, 青木繁伸, 2017, 結合の強度を測る指標としての Log-r の有用性：日・英語の倍グラムデータに基づく MI、LLR などとの比較, 言語資源活用ワークショップ 2016, 国立国語研究所。
5. 藤村逸子, 2016, フランス語の使役文における被使役者の格表示と被動作性の度合, フランス語談話会「他動性と動詞構文」, 東京外国語大学シンポジウム・ワークショップ・パネル
6. Fujimura, Itsuko, 2016, Les sequences "NH + N ou N +NH" : observations dans les grands corpus, NHUMA conférence, ストラスブール (フランス)
7. 藤村逸子, 2016, Caractéristiques de « N1 N2 (épithète) » par rapport à « N1 de N2 »; *Effet domino vs Effet de serre, Fin janvier vs Fin de janvier*, 日本フランス語学会研究促進プログラム「パロールの言語学」大阪
8. Fujimura, Itsuko, Aoki, Shigenobu, 2015, Elaboration of a new score: Log-r for characterizing the types of collocations, Europhras2015, マラガ (スペイン)

〔図書〕(計 2 件)

1. 藤村逸子 2017, *femme médecin の不思議：複合語 < femme + N > の構造に関する日仏語対照* 青木三郎編, 『フランス語学の最前線』第 5 号、ひつじ書房. pp.309-344.
2. Fujimura, Itsuko, Baumert, Nicolas 2014, Le sucre et la saveur sucrée au Japon: l'acculturation sexuée d'une saveur connotée péjorativement, *Le sucre, entre tentations et réglementations*, Archives Nationales du Monde du Travail, pp. 161-176

6. 研究組織

(1) 研究代表者

藤村 逸子 (Fujimura Itsuko)  
名古屋大学・文学研究科 教授  
研究者番号：50229035