様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２８ 年　 ６ 月 ２８ 日現在

機関番号：３２５０６
研究種目：基盤研究(C)（一般）
研究期間：2011～2015
課題番号：２３５２０４７１

研究課題名（和文）語彙情報プロファイリングに基づくフィンランド語の派生要素を含む構文の生産性評価

研究課題名（英文）Corpus-based Lexical Profiling of Modern Finnish Language and its Application to Syntactic Study

研究代表者
千葉　庄寿（Chiba, Shoju）

麗澤大学・外国語学部・教授

研究者番号：７０３３７７２３

交付決定額（研究期間全体）：（直接経費）　　3,800,000円

研究成果の概要（和文）：本研究では，現代フィンランド語の大規模コーパスを用いて形態統語情報データベースを構築し，派生語が含まれる構文の語彙的・構造的多様性を測ると共に，その特徴を抽出・記述する「語彙情報プロファイリング」の手法の開発をおこなった。
構文パターンの特徴を統一的に測る統計的指標を提案することはできなかったが，派生語のもつ生産性とその派生語が含まれる構文の用法・パターンに一定の関係がみられること，派生の生産性を測るためにジャンル比較が有効であることが示された。また，語彙情報プロファイリングの手法を言語教育や他言語の分析に応用する可能性について検討し，専門家を招聘し研究会を開催した。

研究成果の概要（英文）：Lexical profiling technique stands for a framework of evaluating quantitatively the lexical characteristics (i.e. profiles) of different linguistic constructions. It calculates the lexical profiles of a set of data and compares them with an overall lexical picture the whole corpus. Based on the large corpus data of written modern Finnish, this paper tackles the questions of estimating syntactic productivity of a construction on the basis of morphological complexity in that construction. This study argues that there are important correlations observable between the morphological productivity of a derivative and the degree of variety of its syntactic surroundings: The more productive a type of morphological derivation is, the more stable (i.e. the less lexically idiomatic) and more uniform (i.e. more characteristic) behavior the syntax of its constructional environments will show.

研究分野：言語学

キーワード：フィンランド語　その他の外国語　コーパス　派生　統語論　国際情報交流　フィンランド

１．研究開始当初の背景

(1) Aronoff (1976)以来，派生に関する研究において「生産性」は形態論的記述の最も大きな関心事のひとつであった。では，形態論的な生産性の高さ・低さは，その派生要素を含む構文の構造的・語彙的多様性にどのような影響を与えるだろうか。構文の生産性の問題は，Barddal (2008)，Zeldes (2012)などの最近のいくつかの論考で統語論の問題としての議論がようやく始まっているが，本研究の射程である，形態論的派生を伴う構文の生産性に関しては，それが形態論と統語論の接点としての面白さをもつものであるにも関わらず，これまで殆ど論じられてこなかった。統語論の理論的問題をはらむ例として盛んに議論された態(ヴォイス)の構文分析においてさえ，その構文形態の多様性や語彙の出現パターンに関心が払われることは殆どなかった。

(2)本研究が提案する「語彙情報プロファイリング」とは，大規模コーパスとそこからサンプリングし形態統語解析情報を付与したデータベースを用い，構文の用例の集合がもつ構造的・語彙的ヴァリエーションをコーパス全体と比較して評価する，言語形式の特徴抽出と生産性評価のための手法である。語彙情報プロファイリングの枠組みを用いることでフィンランド語の構文ならびに形態統語論の記述精度の向上をめざすとともに，語彙情報プロファイリング技術を他言語にも応用し，これまで量的視点が不足していた構文分析をはじめとするさまざまな言語形式の分析をすすめることで類型論の精緻化への貢献をめざした。

＜引用文献＞

① Aronoff, Mark, *Word Formation in Generative Grammer*, MIT Press (Cambridge, MA, USA), 1976.
② Barddal, Jóhanna, *Productivity: Evidence from Case and Argument Structure in Icelandic*, John Benjamins (Amsterdam, Holland), 2008.
③Zeldes, Amir, *Productivity in Argument Selection: From Morphology to Syntax*, Walter de Gruyter (Berlin, Germany), 2012.

２．研究の目的

(1)本研究は，従来の構文研究で充分に捉えることができなかった「構文の生産性」の評価のための枠組みを整備することを目的として構想された。具体的には，現代フィンランド語の形態論的派生をともなう構文に注目し，当該の形態論的特徴をもつ構文に現れる語彙情報を量的に評価（プロファイリング）することで，その特徴の分析を試みた。複雑な形態論をもつフィンランド語の事例を手がかりとして考察することで，統語分析において見過ごされてきた，構文の形態・語彙パターンの多様性の記述のための基礎的な知見を得，構文の生産性の評価に必要な枠組みを構築することができると考えた。

(2)より具体的には，本研究の活動は大きく 2 つの目的をもつ。まず，第 1 の目的は，本研究が分析対象とする形態論的派生をともなう構文について，①用例中の語彙の出現頻度②語彙の統語的位置および出現環境，ならびに③これらの関連，さらには④コーパス全体の出現状況と構文の用例集合との比較，という観点で構文の用例のもつ語彙的・構造的特徴を抽出するしくみを開発・整備することである。

(3)次に，(2)の技術をもちいて構文の構造的・語彙的多様性を評価する「語彙情報プロファイリング」の手法を開発し，具体的な分析を通じて構文の生産性評価に関する基礎的な知見を得ることが第 2 の目的である。より具体的には，派生を含むフィンランド語の諸構文の分析を通じ，構文の生産性に関する以下のような理論的な問いに答えることをめざす：①形態論的生産性と構文構造および語彙情報パターンの多様性の関係②出現頻度や意味が類似する構文との差異③構文特徴の偏りや特異性の存在の有無。

３．研究の方法

(1)研究前半では，まず構文の語彙的・構造的特徴を抽出するための方法論の開発と分析ツールの整備・開発に取り組んだ。具体的には，フィンランド語の形態統語情報データベースの整備をおこなうとともに，語彙情報プロファイリング情報の統計処理，特に指標化に関して情報収集をおこなうとともに，データベース検索システムおよび語彙情報の分析プログラムの試作，構文の形態統語構造情報の XML による統合的記述のしくみの考察をおこなった。

(2)研究後半では，派生を伴う構文の構造的・語彙的多様性を評価する「語彙情報プロファイリング」の手法に基づくフィンランド語の大規模コーパスから抽出した形態論的派生要素を含む諸構文の分析をおこなうことを試みた。

(3)調査の過程において，本研究が分析対象としているフィンランド語の大規模コーパスのデータが書きことばの特定のジャンル(特に新聞データ)に偏っていることに起因し，分析対象とする構文の語彙パターンを異なるジャンルのコーパスを用いて多角的に検証することが当初開発していた形態統語情報データベースでは難しいことが判明した。そのため，データの検証方法として，平成 26 年度に新たに公開されたフィンランド語の大規模な電子フォーラム(Suomi 24)コーパスを比較データとしてデータベースに取

り込むこととした。そのため，従来の構文データベースの大幅な改編・拡張が必要になったが，結果として文語に偏ったこれまでのデータベースの弱点を補完し，現代フィンランド語の書き言葉のデータの全体像の描出に役立つデータベースが得られたと考える。

４．研究成果

本研究で得られた成果は以下のとおりである。なお，当初計画よりも分析手法の開発が遅れ，結果としてデータの構築に大幅な遅延が生じたことに伴い，研究期間中の研究成果は国際学会での研究発表に重点をおいておこない，専門家との情報交換を積極的におこなうことができた。本研究で構築したデータを用いた研究論文および研究データを応用した成果物を今後も積極的に公開してゆく予定である。

(1)語彙情報プロファイリングをおこなう構文の用例データベースの基本部分について構築を完了し，その分析をすすめた。動詞派生の名詞群の比較分析では(研究発表④)，コーパスに１回限り現れた語(hapax legomena)の数に基づき派生語の生産性を測る場合，派生語が生産的であればあるほど，派生の用法が一貫する傾向がみられることが明らかになった。さらに，動詞派生の名詞の統語的環境において，生産性の高いグループ(-*minen*, -(*i*)*m0*, -(*nn*)*e*)には属格名詞句の修飾をうける比率が相対的に低いこと，またこれらが主語の位置にくる比率が相対的に低いことなど，共通の統語的特徴を持っていることが明らかになった(研究発表③)。

研究当初に想定していたプロファイリングの指標化による統一的な分析については，専門家との情報交換により，指標化ではデータベースから得られる多様な情報を総合的に評価・分析することが困難であるという見通しが得られた。そこで，新たなデータ分析手法として，形態論的生産性のような頻度に基づく統計的指標の利用と並行して可視化(visualization)の手法を用いることを検討し，分析環境の整備をおこなった。結果として計画当初には想定していない大幅な遅延が生じる結果となったが，構文の生産性評価の問題を解決するための方法論的な見通しがある程度得られたので，今後は視覚化を援用した分析・評価手法の開発にさらに注力していきたい。

(2)語彙情報プロファイリングによる多様な観点での語彙の特徴評価に関して，語彙のつくるネットワークという観点から形態論の派生のパラダイムについての新たな知見が得られた。研究発表(②)では，類義の派生辞が必ずしも同一の語根に接続しないこと，また意味的には無関係の派生辞が形式的な理由のあり・なしに関わらず「共生」している例があること，またこれらは生産性の観点から必ずしも生産性が高いとはいえない「中程度」の生産性をもつ派生パターンに多く見られることを指摘した。

(3)語彙情報プロファイリングのデータベースの構築過程で，データベースが依拠しているフィンランド語の大規模書き言葉コーパスが新聞データに偏っていることに対し専門家から問題が指摘された。実際，このことに起因して，フルサイズのコーパスをサンプリングして作成した解析精度の高いコーパスを用いて分析するとサブコーパスが一部小さくなりすぎジャンル間の差異が充分に検出できない問題があることがわかっており，この問題を解決するため，構築したデータベースに新たに異なるジャンルのコーパスデータ(電子フォーラムコーパス)を加えることで現データベースのジャンルの偏りの補正を試みることにした。データベースの形式の修正をおこない，拡充したデータベースを用いて再度取得した用例データと比較した結果，形態論的派生の生産性にジャンル間の違いが明確に観察される事例があることが分かった(研究発表①)。派生のもつ創造的側面の観察がジャンルの影響を受けることは，形態論の分析において多様なジャンルのデータの観察が非常に重要であることを示唆しているといえる。

(4)本研究では，研究の過程で研究手法のさまざまな応用の試みに関する知見が得られた。語彙情報プロファイリングで用いられる語彙の頻度情報は，取りも直さずテキストの語彙多様性を表す指標でもある。頻度情報を用いた指標には，コーパスサイズの影響が少ないとされるものがあり，語彙情報プロファイリングの目指す分析においても重要な観点となっている。

本研究では，その過程において語彙情報を用いた各種統計指標に関する知見を様々な種類のコーパス，フィンランド語以外の言語に応用することをめざしており，この分野でもある程度の成果を得ることができた。現代日本語の大規模な書き言葉コーパス(BCCWJ)と学習者むけ日本語教科書の比較(論文③)はその一例である。さらに，フィンランド語学習者コーパスの専門家と大規模コーパスデータの比較分析について共同研究の準備をすすめ，2014 年には研究者を招聘しコーパスデータを用いた応用言語学的成果物の構築に関する研究会(2014 年 7 月 10 日，言語研究センター研究セミナー，於麗澤大学)を，また学習者コーパスの分析と語彙情報プロファイリングの応用に関する学会講演・研究発表を実施することができた(2014 年 7 月 5 日，第 41 回日本ウラル学会研究大会，於中京大学)。学習者コーパスなど比較的サイズが小さいコーパスデータで語彙情報プロファイリングを用いた分析をおこなう可能性については，研究手法の開発と合わせ今後も

検討を続けていきたい。

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６件）
①千葉庄寿，形態論的生産性を測る指標とコーパスサイズの関係について，東北大学言語学論集，2016，掲載頁未定．
②滝沢直宏，千葉庄寿，コーパス検索の方法，コーパスと日本語教育，2016，pp. 1-34.
③千葉庄寿，曹大峰，井上優，日本語教科書の分析，コーパスと日本語教育，2016，pp. 107-134.
④FUJIMURA，Itsuko，Shoju CHIBA，Mieko OHSO，Lexical and grammatical features of spoken and written Japanese in contrast: Exploring a lexical profiling approach to comparing spoken and written corpora，Proceedings of the 7th GSCP International Conference: Speech and Corpora，2012，pp. 298-301.
⑤千葉庄寿，HTML と XML，ウェブによる情報収集，2011，pp. 177-227.
⑥千葉庄寿，大規模コーパスを用いた用例の典型性評価―大規模コーパスを利用した学習辞書作成のために―，第１回コーパス日本語学ワークショップ予稿集，2012，pp. 185-194.

〔学会発表〕（計９件）
①千葉庄寿，形態論的生産性を測る指標とコーパスサイズの関係について―現代フィンランド語の２種類の大規模コーパスを用いた考察―，第 43 回日本ウラル学会研究発表大会，2016-07-02，東京外国語大学本郷サテライト(東京都・文京区)
②CHIBA，Shoju，The lexical realization pattern and their relationship to productivity: the Finnish symbiotic derivation，17th International Morphology Meeting (IMM17)，Vienna (Austria).
③CHIBA，Shoju，How does morphological productivity facilitate syntactic consistency? A corpus-based study of modern written Finnish，XII International Congress for Finno-Ugric Studies (FU XII)，2015-08-20，Oulu (Finland).
④ CHIBA，Shoju，When derivational morphology meets context: A corpus-based study of Finnish nominalization，42nd International Systemic Functional Congress (ISFC2015)，2015-07-31，Aachen (Germany)
⑤CHIBA，Shoju，Johto-opin leksikaalinen verkosto ja sen visualisointi: alustava tutkimus nykysuomen verbikantaisista johdoksista，XLII Finnish Conference of Linguistics，2015-05-21，Vaasa (Finland).
⑥ CHIBA，Shoju，Evaluating the productivity of Finnish word formation: A quantitative study，第 41 回日本ウラル学会研究大会，2014-07-05，中京大学(愛知県・名古屋市).
⑦千葉庄寿，現代フィンランド語の動詞派生の複雑性と構文パターンの関係について，国語研共同研究プロジェクト「述語構造の意味範疇の普遍性と多様性」研究会，2014-07-27，新潟大学(新潟県・新潟市)
⑧千葉庄寿，大規模コーパスを用いた用例の典型性評価―大規模コーパスを利用した学習辞書作成のために―，第１回コーパス日本語学ワークショップ，2012-03-05，国立国語研究所(東京都・立川市).
⑨千葉庄寿，BCCWJ の量的言語情報の活用―語彙情報のプロファイリングを例に―，『現代日本語書き言葉均衡コーパス』完成記念講演会，2011-08-02，JA 共済ビルカンファレンスホール(東京都・千代田区).

〔その他〕
ホームページ等：
http://www.tsibale.com/publications.html

６．研究組織
(1)研究代表者
千葉　庄寿（CHIBA，Shoju）
麗澤大学・外国語学部・教授
研究者番号：７０３３７７２３