

平成21年5月31日現在

研究種目：若手（B）
 研究期間：2007～2008
 課題番号：19720111
 研究課題名（和文）コーパス分析に基づく認知言語学的構文研究と日本語教育文法への応用
 研究課題名（英文）Corpus-Based Approach to Cognitive Syntax and its Application to Educational Grammar of Japanese
 研究代表者
 李 在鎬（LEE JAE-HO）
 独立行政法人 情報通信研究機構・知識創成コミュニケーション研究センター言語基盤グループ・専攻研究員
 研究者番号：20450695

研究成果の概要：本研究では、コーパス分析的方法で構文研究を行うと同時に、その成果の教育的応用に向けた基礎資料を作成した。調査研究の成果として動詞の制約から予測できない文の使用実態が明らかになり、名詞の意味に基づく文意の予測実験では8割近い正答率を示した。このような事実から、(特に名詞の) 意味情報が明記されたデータベースの必要性が明らかになったため、言語習得研究への活用を想定したタグ付き学習者コーパスを作成し、公開した。また、コーパス内の全データを複数のオプションで柔軟に検索するシステムも開発しており、人文系の研究者にも簡単に使える仕様になっている。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	600,000	0	600,000
2008年度	500,000	150,000	650,000
年度			
年度			
年度			
総計	1,100,000	150,000	1,250,000

研究分野：理論言語学、外国語教育

科研費の分科・細目：言語学・日本語学

キーワード：(1) コーパス (2) 構文文法 (3) 学習者コーパス
 (4) 形態素解析 (5) 具体的文法記述 (6) 日本語教育
 (7) タグ付きコーパス

1. 研究開始当初の背景

どのような形の言語研究であれ、経験科学としての妥当性を満たすためには、実際の言語使用に基づくものでなければならない。また、日本語の研究に限って言うなら、その成果は日本語教育に資するものでなければならない。このような問題意識のもと本研究で

は、認知言語学的構文研究に、コーパス言語学的視点を取り入れること、また研究成果を日本語教育に還元させることを目的とし、2年間において研究を行った。具体的な研究背景として、認知言語学では意味による動機づけを重要視する。また、個々の構文にはそれぞれが他に還元できない意味があると考え

ている。さらに、言語現象の記述の基本姿勢として用法基盤であることが重要とされている。ただ、方法論に関する部分は、十分に確立されていないという課題がある。こうした課題に対して、コーパス言語学的視点を取り入れることで、一貫した方法による用法基盤の構文研究・意味研究の方法論が確立できる。さらに、研究調査の結果が日本語教育分野での応用・活用できるための工夫として、学習者コーパスと専用の検索システムを開発した。

2. 研究の目的

コーパス分析的方法で構文研究を行うと同時に、その成果の教育的応用に向けた基礎資料を作成する。具体的には、1) 自動移動構文「XがYにVする」の全体の意味とその構成要素としての名詞の意味が持つ関係性を明らかにする。2) 研究調査の結果を取り入れたデータベースを構築し、日本語教育分野に資する。

3. 研究の方法

自動移動構文に関する調査は、二段階で行われた。第一段階では、寺村(1982)が挙げている「コトの類型」でリスト化された格パターン(「XがYに」)と動詞の組み合わせを文意の観点から調査した。第二段階では、コーパスデータを用いて、文意の決定に関わる名詞の(意味的)貢献度を調査した。

第一段階の調査手順は以下の通りである。

- (1) 寺村(1982)における「対面表現」「移動表現」「変化表現」に含まれる動詞すべてをリストアップする。
- (2) 『日本語基本動詞用法辞典』を使用してリストアップされた動詞の自動移動構文に関わるすべての具体例を取り出す。
- (3) 取り出した用例の意味的特徴を4名の母語話者にコーディングしてもらう。
- (4) コーディング結果を標準化し、クラスタ分析する。

第二段階では三つのコーパス(新聞データとして読売新聞コーパス、小説データとして新潮文庫100冊コーパス、新聞と小説の中間的データとして新潮新書コーパス)を使用し、「XがYにVする」パターンに生起するすべての用例を抽出し、分析した。以下の手順で調査を行った。

- (1) 三つのコーパスを形態素解析し、パターン抽出の方法で「XがYにVする」パターンに生起するすべての用例を抽出した。
- (2) XとYの名詞の意味分類を「日本語語

彙大系」から半自動でタグ付けした。

- (3) (2)の名詞のタグ付けと平行し、文全体の意味を6人の母語話者にタグ付けしてもらった。
- (4) 決定木分析で、(2)のXとYの意味分類から(3)の母語話者がタグ付けした文全体の意味を予測するタスクを行った。

4. 研究成果

第一ステップの結果として、図1に示す5つのクラスタが確認された。

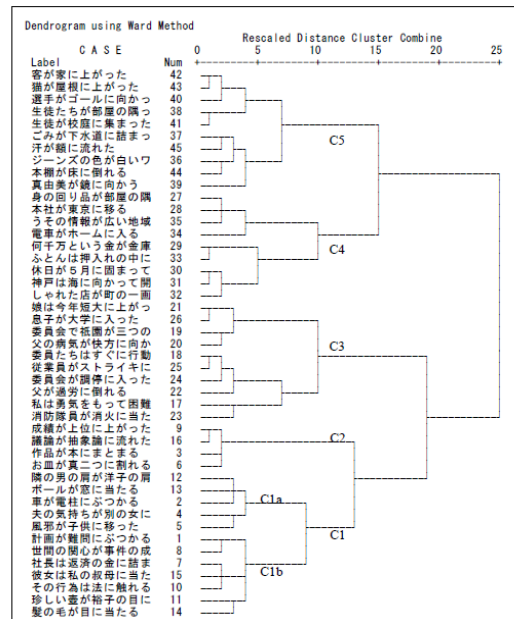


図1 XがYにVする構文の意味クラスタ

図1は、寺村(1982)が言う対面、移動、変化という意味カテゴリをプリミティブに仮定せず、全体の意味と構成要素の意味を総合的に検討し、グループ化したものである。その結果に関して、以下の四点の傾向が見られる。

- (1) 全体の分岐は、動作性を持つ事例(C4、C5)と状態性を持つ事例(C1、C2、C3)に分岐する。
- (2) 動作性を持つ事例は、Xが被動作主として解釈される事例(C4)と、それ以外の事例(C5)に分岐する。なお、C5は必ずしもXが動作主のものとは限らない。
- (3) 状態性を持つ事例は、主としてXは有生物で、自らの意思を伴って何からの状態変化をもたらした事例(C3)と、それ以外の事例(C1、C2)に分岐する。
- (4) C1とC2の分岐は、主としてXの状態変化の結果としてYになったことを表す事例(C2)とそれ以外の事例(C1)に分岐する。

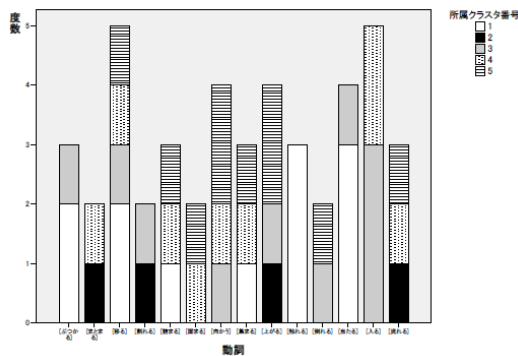


図 2 動詞と意味クラスタのクロス集計

これらの結果を踏まえ、動詞と C1 から C5 の意味クラスタの対応を見るべく、「クラスタ×動詞」でクロス集計をとってみた。図 2 に注目すると、同じ動詞であっても、違う意味クラスタに分類される事例が多数観察される。つまり、同じ動詞の同じパターンであっても、その文全体としての意味は多様であることが明らかになった。これは動詞の意味から単純に文全体の意味を予測することができないことを強く表している。となると、残る可能性として、名詞が文全体に与える意味の問題を検討する必要がある。このことを確認すべく、第二ステップでは、名詞の意味を取り入れたコーパス調査を行った。前節の(第二段階の)(4)のタスクを行った結果、名詞の意味分類から文全体の意味を 78% の正答率で予測できた。このことを踏まえて考えて

みた場合、構文の意味が決定されるためには、名詞の意味に関する情報がなくてはならないことが明らかになった。具体的には、移動構文、対面構文、変化構文の意味的实现には X と Y がもつ図 3 の意味的特徴が大きく貢献していることが明らかになった。

さて、こうした事実を反映したデータベース構築のため、意味タグ付き学習者コーパスとそ

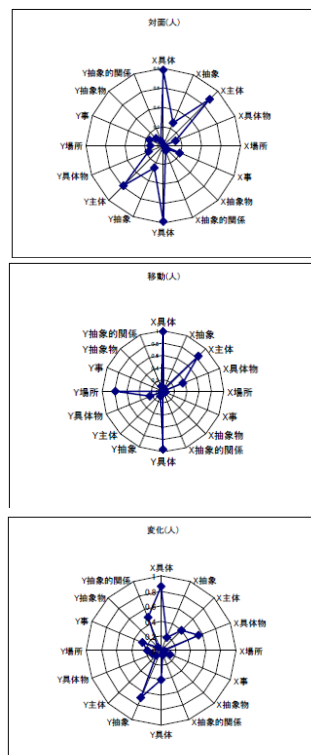


図 3 名詞の貢献度

の検索システムを構築し、公開した。意味タグ付きコーパスとは、OPI に基づくコーパスとして公開された「KY コーパス」に対して半自動で形態素解析情報やエラー情報、意味情報を付与したコーパスである。このコーパスには、データ本体のほか、エクセルで形態素解析済みデータを検索できるソフトウェアの「E-KWIC」も無償で提供している。コーパス本体も利用申請をすれば、無償で提供している。また、本研究の調査結果による直接的な成果ではないが、結果として生まれた成果として 2 点ある。

1. コーパス研究のための文献リスト
2. 算出に基づく語彙リスト

1 はコーパスを用いた先行研究をまとめたリストである。2 は上記のタグ付き学習者コーパスから作成したデータベースで、ウェブから簡単に閲覧やダウンロードができるため、様々な方に利用してもらっている。いずれのデータも <http://www30.atwiki.jp/corpus-ling> において公開されている。

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 8 件)

- ① 李在鎬、移動動詞に対する実験的分析、言葉と認知のメカニズム、87-101、(2008)、査読無
- ② 李在鎬、伊藤健夫、決定木を用いた多義語分析:多義動詞「出る」を例に、認知言語学論文集、Vol 8、55-65、(2008)、査読有
- ③ 李在鎬、用法基盤による言語研究の方法論とその実践に向けて。人環フォーラム、No.22、44-46、(2008)、査読無
- ④ 黒田航、李在鎬、WWW を丸ごとデータにすると何がわかるか?、日本語学、Vol.27 No.2、22-36、(2008)、査読無
- ⑤ 李在鎬、鈴木 幸平、永田 由香、黒田 航、井佐原 均、動詞「流れる」の語形と意味の問題をめぐって、計量国語学、第 26 巻 2 号、64-74、(2007)、査読有
- ⑥ 濱野寛子、李在鎬、助数詞「本」のカテゴリ化をめぐる一考察—統計的手法を用いて、言語学と日本語教育、No.5、73-90、(2007)、査読有
- ⑦ 佐野香織、李在鎬、KH Coder で何ができるか~日本語習得、日本語教育研究利用への示唆~、言語文化と日本語教育、No.33、47-48、(2007)、査読無
- ⑧ 黒田航、李在鎬、Pattern Matching

Analysis (PMA) を用いた日本語の結果構文の共述語分析、小野 尚之(編)結果構文研究の新視点、249-288、(2007)、査読無

[学会発表] (計 12 件)

- ① 李在鎬、コーパス分析と認知言語学、日本認知言語学会、2008.9.14、名古屋大学
- ② Natsuko Nakagawa、Difference between literal and metaphorical usage of verb forms in Japanese、Language、Communication and Cognition: International Conference、2008.7.5、Brighton、UK
- ③ 李在鎬、形態素情報からみた OPI データ、2008 日本語教育学会国際研究大会、2008.7.12、韓国ブサン外国語大学
- ④ 小川典子、コーパス調査による形容詞の連体形と連用形の頻度、2008 日本語教育学会国際研究大会、2008.7.12、韓国ブサン外国語大学
- ⑤ 浅尾仁彦、日本語学習者コーパス検索ツールの開発、言語科学会第 10 回年次大会、2008.7.12、静岡県立大学
- ⑥ 李在鎬、タグ付き日本語学習者コーパスの開発、語処理学会第 14 回年次大会、2008.3.19、東京大学
- ⑦ 李在鎬、決定木を用いた多義語分析:多義動詞、出る」を例に、第 8 回日本認知言語学会、2007.9.17、お茶ノ水女子大学
- ⑧ 李在鎬、助数詞の習得に見られる非対称性、第 24 回日本認知科学会、2007.9.3、成城大学
- ⑨ 李在鎬、初級学習者の名詞使用と 文生成の問題、OPI 国際シンポジウム、2007.8.18、大学コンソーシアム京都
- ⑩ OSAKI Azusa、A Hierarchy of Semantic Constraints in Japanese Fictive Motion、In 10th International Cognitive Linguistics Conference、2007.7.12、Kraków、Poland
- ⑪ 李在鎬、コーパス分析に基づく連体表現の使用調査、日本言語学会第 134 回大会、2007.6.17、麗澤大学
- ⑫ 佐野香織、定住外国人の名詞使用—自然言語処理分析ツールを用いた研究事例—、日本語教育学会、2007.5.26、桜美林大学

[図書] (計 1 件)

- ① 中本敬子、李在鎬、黒田航、ひつじ書房、新しい認知言語学研究法入門、2009(予定)、150-250

[産業財産権]

- 出願状況 (計 0 件)
- 取得状況 (計 0 件)

[その他]

ホームページ:

<http://www30.atwiki.jp/corpus-ling>

ソフトウェア茶豆用検索システム:

<http://www30.atwiki.jp/corpus-ling/pages/51.html>

タグ付き KY コーパスと検索システム:

<http://www30.atwiki.jp/corpus-ling/pages/57.html>

6. 研究組織

(1) 研究代表者

李在鎬 (LEE JAE HO)

独立行政法人 情報通信研究機構・知識創成コミュニケーション研究センター言語基盤グループ・専攻研究員

研究者番号: 20450695

(2) 研究分担者

なし

(3) 連携研究者

なし