

『計量国語学』アーカイブ

ID	KK300602
種別	論文B
タイトル	Web日本語Nグラムコーパス分析に基づく 深層格の偏りの検証
Title	A Corpus-Based Study of Deviation Frequencies in Japanese Deep Cases
著者	松田 真希子
Author	MATSUDA Makiko
掲載号	30巻6号
発行日	2016年9月20日
開始ページ	344
終了ページ	356
著作権者	計量国語学会

論文 B

Web 日本語 N グラムコーパス分析に基づく
深層格の偏りの検証

松田 真希子 (金沢大学)

要旨

日本語の深層格のバリエーションは研究者によって隔たりが大きく、これらの提示根拠がコーパスベースで分析されたものは少ない。本研究では助詞ガ、ヲ、デ、ノについて Web 日本語 N グラムコーパスに深層格情報を付与し、出現頻度を検証した。検証の結果、下記の結果が得られた。助詞ガで最頻度で出現する深層格は非対格自動詞主語 (32% (TYPE; 以下同)) と対象 (27%) であり、動作主体は 8% にすぎなかった。助詞ヲで出現する深層格は対象 (非変化) であり、全体の 8 割を占め、対象 (変化) を合わせると 9 割を超えた。出発点や経路はそれぞれ全体の 0% であった。助詞デで出現する深層格はその他 (51%) が最多で、次いで材料・手段 (29%)、行為の場所 (11%) であった。助詞ノで出現する深層格は限定・修飾 (47%) が多く、所有は 3% にすぎなかった。本研究によって導かれる結果には、Web コーパスの特性の影響が否めないが、深層格の偏りをコーパスベースで検証したことは言語研究上意義があると言える。

キーワード: 深層格, Web 日本語 N グラムコーパス, ガ格, ヲ格, デ格, ノ格

1. はじめに

格は名詞・代名詞が文中で他の語とどのような関係にあるかを示す文法範疇である。日本語において、通常、格はガ、ヲ、ニなどの形式としての表層格と、「動作主」や「対象」等の意味的な役割を担う深層格に分けられている (仁田 1993)。しかし、深層格の定義、命名、類型提示は研究者によって隔たりが大きい。例えば、松田他 (2014) は、6 人の研究者による二格深層格の類型を言及例から整理したところ、全ての文献で共通に言及されている深層格は「時間」「場所」「対象」の三つであることを述べている (表 1)¹。

山梨 (2014) が深層格 (意味役割) は「完全性」「個別性」「唯一性」「独立性」「自然性」といった基準が前提となっていると述べているように、深層格の定義については、基本的に意味や言語的特性において検討されている。そのため、深層格のコーパスベースの量的検証は荻野・植田他 (2004, 2005)、松田他 (2014) など限られている。

1 このでの“○”は各研究者がそれぞれの例に言及し何らかの深層格名を与えていることを示している。同様の現象にどのような深層格名を付与するかは研究者によって異なっている。特に「二格」の「対象」は範囲も広く、同一例に対する命名のバリエーションが大きい。例えば、庵他 (2000) では、「父に手紙をあげる」の「に」に「受け手」という名前を挙げており、石綿 (1999) では「動作のゆきつく相手」という名前を挙げている。

表 1：二格深層格の種類の隔たり（松田他 2014 を改編）

松田他 (2014) 深層格	例	城田 (2002)	鈴木 (1978)	奥田 (1983a)	石綿 (1999)	高橋 (2005)	庵他 (2000)
時間	8 時に起きる	○	○	○	○	○	○
場所	公園に現れる			○	○		
	ハワイにいる	○	○	○	○	○	○
	東京に行く / 着く	○	○	○	○	○	○
結果	息子を医者にする			○	○		○
	どろどろに溶ける			○	○	○	○
	コの字型にならべる			○	○		
	医者になる		○	○	○	○	
対象	実験に成功する	○				○	
	A は B にまさる	○			○		
	対応に怒る	○		○	○	○	
	説明に困る, 彼にほれる	○		○	○	○	
	駅に近い, 父に似る	○	○			○	
	父に手紙をあげる	○	○	○	○		○
	服にくっつく	○	○	○	○	○	
	お母さんに甘える	○	○	○			
動作主	太郎に殴られる	○			○	○	
	私にできること	○			○	○	
目的	映画を見に行く		○	○	○	○	
役割	貿易を外交の手段に用いる			○		○	
頻度	三年に一回						○
副詞化	元気に歩く, お気軽に申し付け下さい			○			
複合辞	環境について語る					○	
	法律に基づく表示		○				
起点	太郎にもらう			○			○
その他	真犯人にちがいない, 口に出す, 役にたつ			○	○		

荻野・植田他（2004, 2005）では動詞が必要とする格助詞の特定の関係（動詞の結合価）に関する調査を行った。大量の動詞の実例データから異なり 12,400 の動詞概念を収集し、どの格助詞の組み合わせが多く出現するかを調査すると共に、どの深層格が多く出現するかについて検討した。その結果、「物を買う」のように「対象格」を取る動詞群が約 55%, 「が」格をとる現象動詞の群が約 18.5%, GOAL の「に」格を採る動詞が 16.9% で、「が, を, に」のいずれかを第一格にする動詞で全体の 91% になることを述べている。しかし動詞にどの格が出現するかに焦点化されているため、助詞ごとにどのような深層格

が出現しているかを理解することはできない。

松田他（2014）は二格を取る動詞句を BCCWJ（国立国語研究所 2011）、京都大学テキストコーパス（黒橋・河原 1997）、Web N グラムコーパス（工藤・賀沢 2007）から二万例ずつ収集し、深層格情報を人手で付与した後、深層格の頻度を比較・分析した。その結果、「N に渡す」「服にくっつく」のような「対象」の深層格がどのコーパスでも最も多く 4 割程度出現すること、時間や場所に関する深層格は合わせて 1 割程度であることを報告している。しかし、ガ、ヲ、デ、ノについては報告されていない。

そこで本稿では松田他（2014）が二格の深層格情報を付与した Web 日本語 N グラムコーパスに、助詞ガ、ヲ、デ、ノ²の四語の深層格情報を付与し、その深層格ごとの出現頻度を調査し、その結果と先行研究で示されてきた深層格の記述との異同について検討することにする。

2. コーパスデザイン

深層格情報を付与したコーパスは Web 日本語 N グラムコーパス（工藤・賀沢 2007）である。このコーパスは Web から抽出した約 200 億文（約 2550 億単語）の日本語データから作成した n-gram データ（1～7gram）を公開したものである。

本研究ではこのコーパスの中で、単語 7gram（異なり 7-gram 数：570,204,252）のデータを用い、格助詞が単語の 4 番目に来るように抽出し、頻度順に 2 万例収集した。分析対象データの概要を表 2、図 1 に示す。

表 2：分析対象データ概要

	TYPE	TOKEN
が	20,010	700,623,154
を	19,995	1,162,519,941
で	19,995	526,836,391
の	19,914	1,620,508,724

次に格助詞のガ、ヲ、デ、ノの 4 つの助詞について深層格のタグセットを作成し、人手でアノテーションを行った。深層格のタグセットについては、庵他（2000）に基づいて選定し、研究チーム内の言語学の専門家から助言を得る形で決定した。コーパスに付与した助詞ごとの深層格タグリストは結果とともに記す。

2 助詞ノは主として名詞と名詞の関係規定を行う助詞のため、ガ、ヲ、ニ、デなどの述語との関係規定を行う表層格とは同列に並べることはできない。しかし本研究では深層格として意味的な広がり大きい助詞について調べることを目的としたため、ノを調査に含めることとした。

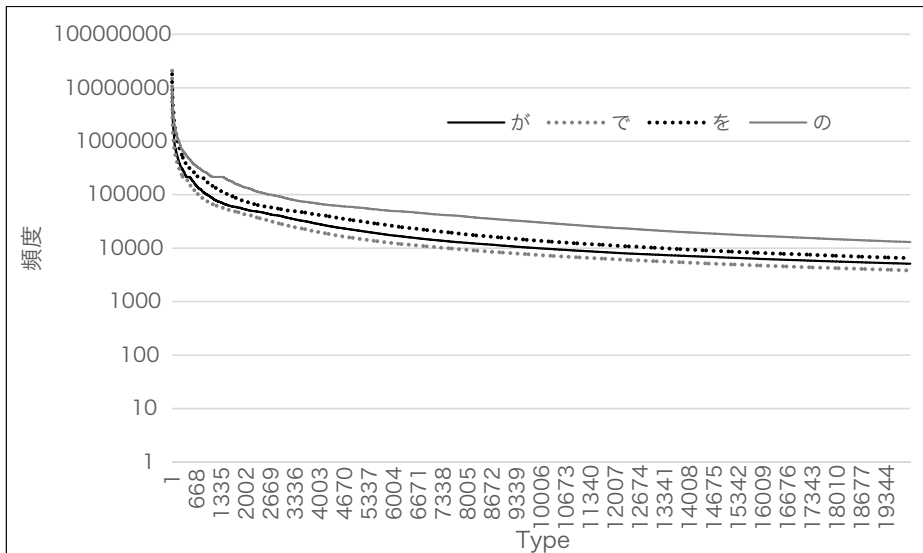


図 1: 分析対象データ概要 (格助詞ごとの 7gram データ出現頻度)

この深層格 Web コーパスを構築する上での問題点として以下の 4 点がある。

- (1) 最長が単語 7 gram のため、短すぎて深層格情報が付与できないデータが存在すること。
- (2) 頻度でソートした Web コーパスのため、Web で高頻度に出現する話題に偏っていること。
- (3) 業者³に委託して付与したデータであるが、二格以外は agreement をとっていないため、アノテーションの誤りも含まれていること。
- (4) 上位 20,000 で切っているの、下位のデータは分析できていないこと。

分析にあたっては深層格ごとの TYPE (異なり数) と TOKEN (延べ数) の算出、前後に出現する単語の傾向分析を行った。

3. 結果と考察

3.1 助詞ガの深層格

最頻度で出現した深層格は「～がある」「～が届く」のような非対格自動詞主語 (TYPE32%) と対象 (TYPE27%) であり、動作主は他動詞主語と非能格自動詞主語の場合を合計しても TYPE8% にすぎなかった (表 3)。さらに、動作主でマークされているものの多くは [1] [2] に見られるように、記事の見出しや連体修飾節内におけるガであった。

3 アノテーション作業は日本システムアプリケーション (JSA) に委託し、日本語学の知識のあるアノテーターが付与した。

[1] リア・ディゾンが応援! (76,447 例)

[2] AllAbout 社が対談する理由 (74,053 例)

表 3: 助詞ガの深層格と出現頻度の偏り

ID	深層格	頻度上位の出現例	TYPE	TOKEN
44	非対格自動詞主語 (対象)	まだ、評価がありません れたメールが届きます.	6,727 (32%)	224,579,646 (32%)
42	対象 (動詞以外)	簡単カスタマイズが可能! この回答内容が不快なら	5,303 (27%)	202,945,235 (32%)
40	その他	ような気がする. </S>	3,978 (20%)	97,402,298 (14%)
43	受動態主語	, アイコンが表示され	1,161 (7%)	72,994,198 (10%)
×	判定不能	ページの更新が定期的に	1,388 (7%)	52,558,153 (8%)
41	他動詞主語 (動作主 体)	中 0 人が役に立つと評価 落札者が送料を負担	1,101 (6%)	44,012,119 (6%)
45	非能格自動詞主語 (動作主体)	口コミ投稿者が宿泊した BS ファンが集まる「掲示板」	350 (2%)	6,082,565 (1%)
47	使役のガ格	ショップ管理者が確認させ	2 (0%)	48,940 (0%)
計			20,010	700,623,154

石綿 (1999:132-133) では、ガの深層格は (1) 動作の主となるもの (A) (例: チームが優勝した) (2) 動作・状態の主となるもの (E) (例: 私は孤独を感じた) (3) 動作・状態の主となるもの (O) (例: 色が変わる) の三つが基本的な用法で、「わかる」「できる」などの動作・状態の目的・対象となるもの (O) (例: 数学がわかる) はその次に記述されている。そして A の動詞は極めて多いと述べている。

コーパスの特性上の偏りがあるとはいえ、ガの深層格として第一に記述されることの多い「動作主 (石綿の A)」の出現頻度が 1 割にも満たず、対象の用法 (石綿の O) が全体の 6 割を占めたのはアンバランスな結果といえる。その理由として最も考えられるのは、ガでマークする必然性の有無の問題である。ガが動作主を表すことは確かだが、実際にはガで動作主を示すことができる場合であっても、ガが省略される可能性が高いこと、動作主体を助詞ガではなく助詞ハでマークすることが多いことなどが理由として考えられる。それに対し、対象や非対格主語としてのガ格は動作主に比べガを明示的に使用する必要性が高いため出現している可能性がある。尾上 (2014) はガ格に立つ名詞項の共通点が「事態認識の中核項」であると述べているが、その前提に立つ場合、動作主ではなく対象や直接目的語を事態認識の中核項として明示することが多いのではないだろうか。

また、TYPE より TOKEN の比率が上回ったものに対象 (27% → 32%)、受動態主語

(7% → 10%) があった。これらについては、「カスタマイズが可能」「内容が不快」のような「A が B」に固定された表現が多いことが窺える。単語 7gram で TOKEN 比率が増加しているのであれば、単語 3gram では更に TOKEN が増加することが予想される。

3.2 助詞ヲの深層格

集計の結果、最頻度で出現する深層格は対象（非変化）であり、全体の 8 割を占めた（表 4）。そして対象（変化）を合わせると 9 割であった。出発点や経路はそれぞれ全体の 0% であった。

この結果は従来の日本語学で指摘されていることをコーパスベースで追認できたといえる。奥田（1983b:22）は「道を歩く」「空を飛ぶ」のような「状況的な結びつき」を表す連語は動詞が自動詞でなければならないという語彙的な制限に強く縛られているため、使用頻度は少なく、様々なタイプの連語に展開していないことを既に指摘している。

表 4：助詞ヲの深層格と出現頻度の偏り

ID	深層格	頻度上位の出現例	TYPE	TOKEN
24	対象 (非変化)	<S>[一覧を見る]</S>	15,552 (78%)	944,935,205 (81%)
23	対象 (変化)	<S> 無料ブログを簡単作成	2,387 (12%)	108,425,149 (9%)
×	判別不能	この写真素材をライトボックスへ	1,563 (8%)	93,534,436 (8%)
20	その他	にうれしいプレゼントをはじめ楽しい	395 (2%)	11,326,090 (1%)
22	経路	国の国境を越えて感染	11 (0%)	439,048 (0%)
26	限度	スレッドが 1000 を超えました	72 (0%)	3,516,731 (0%)
25	使役主格	奇跡～世界を驚かせた	4 (0%)	181,942 (0%)
21	出発点	は、リングを去るや愛する	11 (0%)	161,340 (0%)
合計			19,995	1,162,519,941

その一方、「動きが対象を目指すといったことの代表・典型は、主体からの物理的な働きかけを受けるものである」（仁田 2009:83）や、「物にたいするはたらきかけをあらわす連語は基礎的な領域をなしていて、を格の名詞と他動詞のくみあわせの研究は、ここから出発しなければならない」（奥田 1983b:24）という記述とは異なり、対象の主要な用法として考えられてきた主体動作・対象変化は全体の 1 割程度にとどまる結果となった。8 割は [3] - [6] にあるような主体動作・対象非変化であった⁴。こうした深層格の主要例とコーパス上の最頻度用法とのずれはヲ格でも確認できる。

- [3] <S>[一覧を見る]</S> (17,901,860 例)
 [4] はインラインフレームを使用して (13,993,691 例)
 [5] て, 注文を確認または変更 (9,498,373 例)
 [6] にだけ表示を許可する </S> (5,733,192 例)

またヲについては、ガと異なり、TYPE より TOKEN のほうが目立って増加する深層格は見られなかった。

3.3 助詞デの深層格

助詞デにおいて最頻度で出現する用法は「その他」で、「無料でお届け」のように述語の様態規定にかかわる用法が、全体の 5 割を占めた (表 5)。材料手段は 29%、行為の場所は 11% である。

表 5：助詞デの深層格と出現頻度の偏り

ID	深層格	頻度上位の出現例	TYPE	TOKEN
50	その他	配送料無料でお届けし	10,181 (51%)	251,403,105 (48%)
52	材料手段	は, プルダウンで「顔選択	5,739 (29%)	185,918,571 (35%)
51	行為の場所	商品ページ等で詳細をご	2,224 (11%)	56,090,817 (11%)
53	原因	<S> 円安で ebay 買う	794 (4%)	10,357,505 (2%)
54	範囲	7 日間集中で結果を出	405 (2%)	10,633,225 (2%)
56	主体	<S>HMVJapan では音楽	362 (2%)	6,465,222 (1%)
55	まとめり	疑問, みんなで解決!	161 (1%)	3,721,155 (1%)
×	判定不能	お久しぶりでーす	129 (1%)	2,246,791 (0%)
合計			19,995	526,836,391

次に「その他」を取り除いたうえで、深層格としての用法だけでみると、最も高頻度で出現したものは材料・手段 (TYPE68%) で、「行為の場所」の TYPE は全体の 28% であった。また「その他」には「無料でお届け」のように、材料・手段との境界用法も多くみら

4 対象非変化には「削除用パスワードを入力して」「ページにファイルをアップロード」「<S> この記事をクリップ!</S>」のような空間的な位置変化動詞も含まれている。これらは変化と非変化の中間にあたるため、これらを対象変化に含める場合、比率が上がる可能性がある。

れた。また、「材料・手段」は TYPE より TOKEN の比率が 6% 上昇していた。固定表現が多いといえる。

助詞デの用法として「行為の場所」を第一に挙げているものはいくつかみられる（石綿 1999, 杉本 2014）。例えば石綿（1999:139）は「行為・イベントの場所」「範囲」「限定／基準」「時間」「道具・手段」の順で述べられている。しかし実際のコーパス上では「材料・手段」が最も多く、「行為の場所」の比率は 1 割程度であった。ここでも第一に示されている用法と高頻度用法とのずれが確認される。

3.4 助詞ノの深層格

最頻度で出現した深層格は限定・修飾（TYPE47%）であった（表 6）。所有のノは TYPE, TOKEN ともに全体の 3% にすぎなかった。同格のノもほぼ 0% であった。『日本語大辞典』（梅棹他 1995）では①主語、②所有・所在・所属、③種類・属性・材料・数量④動作の対象、⑤基準となるもの、⑥同格の順で示されている。本研究では③が最も多く出現し、①、②は非常に少なかった。ここにも提示順と高頻度用法とのずれが見られた。

ヲ格、相対的關係については TYPE より TOKEN のほうが 3～4% 上昇している。これらについては定型表現が多い可能性がある。

表 6：助詞ノの深層格と出現頻度の偏り

番号	深層格	頻度上位の出現例	TYPE	TOKEN
36	限定・修飾	国内最大級の無料ブログサービスです	9,455 (47%)	739,559,921 (46%)
30	その他	～ていたのですが	3,267 (16%)	190,636,951 (12%)
38	ヲ格	この商品のレビューを見る	1,870 (9%)	191,593,792 (12%)
34	相対的關係	<S> このページのトップへ </S>	1,442 (7%)	17,282,586 (11%)
391	複合辞+の	この記事へのトラックバック URL	1,358 (7%)	110,543,193 (7%)
37	主格	クレジットカード決済の可能なショップ	676 (3%)	45,520,617 (3%)
31	所有	トラックをあなたのプレイリストに追加	548 (3%)	56,070,551 (3%)
35	時間	過去一時間の間にもっとも	525 (3%)	34,057,851 (2%)
×	判定不能	日本ナンバー 1 の本格的な	257 (1%)	21,941,712 (1%)
39	「による」	店員さんのコメント待ちです	221 (1%)	23,008,180 (1%)

33	場所	はこのサーバを使って	205 (1%)	14,955,326 (1%)
392	a-object ⁵	登録日の新しい順	90 (0%)	17,564,165 (1%)
32	同格	内閣総理大臣の安倍晋三	78 (0%)	4,773,879 (0%)
			19,914	1,620,508,724

4. 深層格コーパスに出現する特徴語からみた WEB コーパスの特徴

前節ではコーパスの頻度情報から深層格の特徴分析を行った。本節では異なりデータにおける特徴語から深層格の特徴について検討する。表7は深層格ごとの特徴語抽出結果を示している。分析にあたっては KH-Coder⁶ (Ver. 2.b.32c) を使用した。Jaccard 係数⁷が 0.1 以上を強い特徴語とみなした。上位 5 語を表7に示す。

抽出の結果、強い特徴語は「ガ: 使役」(例: キュンキュンワクワク, ピコピコ), 「ガ: 受身主語」(例: 場合), 「デ: 範囲」(例: 間, 週), 「ヲ: 出発点」(例: 去る, 離れる), 「ヲ: 経路」(例: 越える, 巡る, 経由), 「ヲ: 使役主語」(例: 力士, 昇進), 「ノ: 同格」(例: 皆様), 「ノ: 対関係」(例: ページ, 前), 「ノ: 時間」(例: 時), 「ノ: による」(例: 規定, 都合), 「ノ: 複合辞」(例: から), 「ノ: a-object」(例: 順, 新しい) などが見られた。

全体傾向として出現頻度が低い深層格は特徴度の高い語が現れたが、出現頻度の高い深層格の特徴語には高い Jaccard 係数の語は見られなかった。また語の傾向からショッピング, ブログ, ビジネス, スポーツ, 交通関連の語が多くみられた。頻度の低い深層格については、かなり語の特徴が明確に表れていた。「デ: まとまり⁸」におけるファミリー関連の語彙, ヲ経路における巡礼関連の語彙などである。これらについては深層格の異なりが WEB コーパスにおけるトピックの特性を反映しているといえる。

5 a-object とは「<S> クチコミ数の 多い順 </S>」「<S> 登録日の新しい順で」など N1 の N2 の N1 が形容詞の対象を表すものである。

6 KH-Coder とはテキスト型 (文章型) データを統計的に分析するためのフリーソフトウェアである。
<http://khc.sourceforge.net/>

7 Jaccard 係数とは 2 つの集合の類似度を表す 1 つの数値表現である。集合 X と Y の共通要素数を少なくとも一方にある要素の総数で割ったもので、本研究の場合、値が 1 に近いほど特徴度が高いとされる。

8 デの「まとまり」とは「家族ででかける」「疑問, みんなで解決!」「三人でデート」「一人で悩む」のような動作の担い手の数的な規模やまとまりを表す。

表 7： 深層格ごとの特徴語 (Jaccard 係数)

ガ： 他動詞主語		ガ： 対象		ガ： 受身主語		ガ： 使役	
運営	.073	必要	.063	場合	.136	キュンキュン ワクワク	.333
語る	.035	出来る	.053	必要	.032	ピコピコ	.333
株式会社	.035	可能	.049	商品	.030	使役	.250
人	.033	方	.043	間	.029	シヨップ	.007
提供	.033	多い	.042	時	.027	管理	.007
デ： 主体		デ： まとまり		デ： 範囲		デ： 原因	
当店	.081	親子	.061	間	.132	都合	.029
サイト	.068	家族	.060	週	.106	対応	.026
弊社	.058	作る	.045	發送	.078	安心	.019
一切	.057	悩む	.034	年	.073	今日	.018
負担	.043	共有	.031	後	.063	問題	.018
デ： 材料手段		デ： 場所		ヲ： 出発点		ヲ： 経路	
メール	.056	サイト	.050	去る	.154	越える	.167
検索	.046	上	.046	離れる	.133	巡る	.111
電話	.038	ページ	.034	離陸	.083	経由	.111
情報	.029	中	.030	半	.083	ミツ	.083
ネット	.024	ここ	.027	シャルル	.083	札所	.083
ヲ： 対象変化		ヲ： 対象非変化		ヲ： 使役主語		ニ： 場所	
書く	.061	情報	.051	力士	.200	中	.060
作る	.035	クリック	.041	昇進	.167	行く	.050
変更	.035	見る	.036	使役	.167	ページ	.034
作成	.033	表示	.035	主語	.143	下	.032
公開	.031	商品	.034	驚く	.083	店	.028
ノ： その他		ノ： 所有		ノ： 同格		ノ： 場所	
方	.102	自分	.084	皆様	.114	全国	.076
思う	.053	日記	.065	カウンセラー	.047	近く	.065
上	.030	声	.054	白石	.038	ホテル	.056
確認	.015	ブログ	.047	義理	.038	周辺	.052
希望	.012	商標	.040	皆さん	.038	地域	.051
ノ： 相対的關係		ノ： 時間		ノ： 限定・修飾		ノ： 主格	
ページ	.117	時	.157	情報	.064	お客様	.034
前	.108	月	.073	商品	.052	間	.025
次	.083	現在	.060	場合	.030	判断	.025
以上	.081	夏	.058	サイト	.024	負担	.025
後	.078	間	.058	中	.023	高い	.024

ノ：ヲ格		ノ：による		ノ：複合辞の		ノ：a-object	
情報	.062	規定	.172	から	.118	順	.730
案内	.049	都合	.120	まで	.064	新しい	.149
整備	.042	項	.118	問い合わせ	.038	安い	.134
個人	.038	責任	.084	リンク	.034	古い	.112
商品	.036	お客様	.059	注文	.030	価格	.069

5. まとめ

本研究では Web N グラムコーパスに人手で付与した深層格情報に基づき、助詞ガ、ヲ、デ、ノの深層格の偏り情報と典型用法とのずれについて明らかにした。本研究によって導かれる結果には、Web コーパスの特性の影響が否めないが、これまで明らかにされていなかった深層格の偏りをコーパスベースで明らかにし、典型的として提示されてきた深層格が必ずしも量的に優位ではないことを示したことは言語研究上の意義があると考えられる。

また、TYPE より TOKEN の比率が高い深層格は、材料手段のデ、対象と受身のガ、対象のヲ、相対的關係のノなどがあった。これらの TOKEN の比率が高い深層格については日本語のコロケーション教育への貢献可能性が窺える。特にガについては、日本語学習では「A を B」「A で B」「A の B」のコロケーション学習は盛んだが、「カスタマイズが可能」のような「A が B」タイプのコロケーション学習は相対的に少ない。そのため「A が B」の高頻度コロケーションを抽出・リスト化することでコロケーション学習が促進される可能性がある。

本コーパスは頻度上位二万例から採取しているため、頻度の低い用例もすべて対象にした場合、深層格の頻度の内訳が変わる可能性も残されている。同時に Web コーパスであるため、実際の話し言葉も含めた言語コーパスでも同様の結果が得られるか検証の必要がある。また、付与した深層格情報の妥当性も高める必要がある。併せて今後の課題とした。

謝辞

本研究は JSPS 科学研究費補助金（科研費）23320105 の助成を受けたものです。

文献

- 庵功雄, 松岡弘, 中西久実子, 山田敏弘, 高梨信乃 (2000) 『初級を教える人のための日本語文法ハンドブック』スリーエーネットワーク.
- 石綿敏雄 (1999) 『現代言語理論と格』ひつじ書房.
- 梅棹忠夫, 金田一春彦, 坂倉篤義, 日野原重明監修 (1995) 『日本語大辞典 第二版』小学館.
- 荻野孝野, 小林正博, 井佐原均 (2003) 『日本語動詞の結合価』三省堂.
- 荻野孝野, 植田禎子, 小林正博 (2004) 「動詞結合価の基本パターンの選定について」『言語処理学会第 10 回年次大会発表論文集』A5-4.
- 荻野孝野, 植田禎子, 小林正博, 井佐原均 (2005) 「コーパスデータに基づく格助詞組み合わせレベルにおける結合価の実態と同音異表記判定での利用」『自然言語処理』

Vol.12, No. 4, 21-54.

- 奥田靖雄 (1983a) 「に格の名詞と動詞のくみあわせ」 言語学研究会編『日本語文法・連語論 (資料編)』 281-323. むぎ書房.
- 奥田靖雄 (1983b) 「を格の名詞と動詞のくみあわせ」 言語学研究会編『日本語文法・連語論 (資料編)』 22-149. むぎ書房.
- 尾上圭介 (2014) 「主語」 日本語文法学会編『日本語文法事典』 167-270. 大修館書店.
- 工藤拓, 賀沢秀人 (2007) 『Web 日本語 N グラム 第 1 版』 言語資源協会.
- 黒橋禎夫, 河原大輔 (1997) 「京都大学テキストコーパス・プロジェクト」『言語処理学会 第 3 回年次大会』 115-118.
- 国立国語研究所 (2011) 『現代日本語書き言葉均衡コーパス』 国立国語研究所.
- 城田俊 (2002) 『日本語形態論』 ひつじ書房.
- 杉本武 (2014) 「デ」 日本語文法学会編『日本語文法事典』 412-413. 大修館書店.
- 鈴木重幸 (1978) 『日本語文法・形態論』 むぎ書房.
- 高橋太郎 (2005) 『日本語の文法』 ひつじ書房.
- 仁田義雄 (1993) 「日本語の格を求めて」 仁田義雄 (編) 『日本語の格をめぐって』 1-38. くろしお出版.
- 仁田義雄 (2009) 『仁田義雄日本語文法著作選 1 日本語の文法カテゴリをめぐって』 ひつじ書房.
- 松田真希子, 森篤嗣, 川村よし子, 庵功雄, 山本和英, 山口昌也 (2014) 「ニ格深層格の定量的分析」『言語処理学会 第 20 回年次大会発表論文集』 516-519.
- 山梨正明 (2014) 「意味役割」 日本語文法学会編『日本語文法事典』 30-31. 大修館書店.

(2016 年 3 月 31 日受付・2016 年 6 月 26 日改稿受付)

A Corpus-Based Study of Deviation Frequencies in Japanese Deep Cases

MATSUDA Makiko (Kanazawa University)

Abstract:

Japanese linguists have proposed a variety of Japanese deep cases from the semantic perspective. However, few people have investigated the adequacy of these cases using corpus-based analyses. This study investigates the frequency of four particles (*Ga*, *Wo*, *De*, *No*) by annotating the deep cases of these particles using the Japanese web n-gram corpus (7 gram). Results indicate that the most frequently appearing deep cases of *Ga* are “unaccusative-intransitive” (32% in type) and “objective” (type = 27%). “Agent” did not appear much (type = 8%), although they are considered one of the prototypical deep cases of *Ga*. “Object (act-on)” cases appear the most frequently in *Wo* (type = 80%) and they become more than 90% with “object (act-cause) cases. “Starting point” and “route” cases did not appear at all (type = 0%), despite their semantic and grammatical uniqueness. The most frequently appearing deep cases of *De* were “others” (type = 51%) and “means and material” (type = 29%). “Place” cases appeared only 11% though they are considered a prototypical deep case. “Limitation and Modification” is the deep case that most frequently appeared in *No* (type = 47%), while prototypical deep case “Possessive” is not frequent (type = 3%). Although the Web influences these results, they may nonetheless provide useful insights for the study of Japanese deep cases.

Keywords: deep case, Japanese web n-gram corpus, *Ga*, *Wo*, *De*, *No*