

# 動詞結合価の基本パターンの選定について

荻野孝野<sup>1,2</sup>、植田禎子<sup>1</sup>、小林正博<sup>1</sup>

{ogino, y-ueda, kobayasi}@jsa.co.jp

1 日本システムアプリケーション

2 神戸大学自然科学研究科

## 1. はじめに

筆者らは、延べ約156,000文から抽出した約12,400概念の動詞で作成した結合価データ[1]を対象に結合価パターンの研究を行なっている。動詞の結合価とは、それぞれの動詞が必要とする格関係である。例えば「行く」では、「{人}が{場所}に行く」というように、必要な格助詞と格助詞がとる体言の意味的傾向が決まってくる。

今回は、結合価のうち、格助詞の組み合わせレベルに注目し、読みは同じだが概念や表記が異なる動詞898概念を対象に、基本格パターンへの集約を試みた。ここで基本格パターンとは、その動詞を特定づけるために必要となる格(必須格)の組み合わせパターンである。

動詞には特有の格助詞の組み合わせが決まっているとしても、表現時には省略される格もあり、实例はさまざまな様相を提示する。これらの多様なパターンから、基本格パターンを設定する過程と問題点などを報告する。

## 2. 本研究の目的

各動詞について格助詞は、いつでも実データの中にその動詞に必要な格助詞が全部出現するとは限らない。また、連体修飾句においては、本来連体句の中の動詞にかかる体言が連体句の修飾先となり、表現上は格関係は表示されていない。自然言語解析をするにあたって、意味的な解析まで範囲を広げたとき、省略された格は何であるか、連体句の修飾先の体言と連体句の中の動詞の格関係は何なのかなどの情報が必要になる。

そこで筆者らは、格関係を整備した関係で見ることができる「日本語動詞の結合価」[1]データを対象に、動詞ごとの格助詞の基本格パターンを設定することを試みている。「日本語動詞の結合価」[1]は、EDRコーパス[2]内の実例文を基に結合価データとしてまとめたものである。

ここでは、まず、元データである結合価データ[1]の異なり約12,400の動詞概念から、読みと表記は同じだが概念が異なるグループ(いわゆる同音異義語)と、同じ読みで複数の表記があるグループ(いわゆる同音異表記)に対象データを限定して分析を行なった。

## 3. 作業手順とデータ

### 3.1 作業データの設定

#### 3.1.1 元データとなった結合価データ

「日本語動詞の結合価」[1]を構成している結合価データを元データとして用いて分析した。これは、表1に示すように、共起辞書で「同じ表記+同じ概念」を単位として、「同じ表記+同じ概念」にかかる格関係に該当する要素を取り出し、下記のように1ファイルにまとめたものである。

表1 結合価データの部分

食べる	タベル	3bc6f0	食物をとる								例文
は	が	を	に	へ	から	より	まで	で	と	その他	
	が(牛(名詞))	(エサ(名詞))									牛がうまそうに[[食べ]]ているエサは、シラカバの木のこま切れ。
		を(サンドイッチ(名詞))	は(昼食(名詞))								あまりおなががすいていなかったので、昼食は、軽くサンドイッチを[[食べ]]た。

#### 3.1.2 例文つき結合価データから格助詞組み合わせの抽出

動詞の結合価データの1ファイル(表1相当)ずつを対象として、その中で格助詞の組み合わせが同じものを1グループとして、「読み、表記、概念ID、概念説明、格助詞組み合わせ、その頻度」の形式に整えたものを作成した(表3と同じ形式)。

#### 3.1.3 作業対象データの絞り込み

作業対象データとなる「日本語動詞の結合価」[1]データから、同じ読みを出発として異なる表記のもの、もしくは異なる概念のものでかつコーパスに出現した頻度が一定以上のものを対象とし、作業データの絞り込みを行った。

### 同音異義語および同語異表記リストの作成

3.1.1に含まれる概念単位から、ヘッダー部(読み、表記、概念ID、概念説明)において、「読み」が同じものを抽出して、それらを1グループとして出力したものが表2である。

表2 読みが同じもののグループ

頻度	よみ	表記	語幹	概念ID	概念説明	調査対象行
240	ア・ウ	会う	会	3ce8ca	会う	*
68	ア・ウ	合う	合	1fa2c9	ぴったり一致する	*
34	ア・ウ	遭う	遭	103c7e	思いがけない人や事件と出会う	*
26	ア・ウ	合う	合	3ce55d	きちんと一致する	*
15	ア・ウ	合う	合	3ce9d3	釣り合いがとれる	*
7	ア・ウ	合う	合	1fa2cc	互いに?する	*
5	ア・ウ	合う	合	1fa2ca	正しいものに合う	*
1	ア・ウ	合う	合	0eb828	(二つ以上のものが)一つに合わさる	
	ア・ウ	合う	合	1fa2cb	引き合う	
	ア・ウ	合う	合	1fa2c7	(二つ以上のものが)一つに集まる	
	ア・ウ	合う	合	1fa2cd	~して一緒になる	

表2を対象に、読みでみてコーパス出現頻度が5以上ある概念が3つ以上あるときは頻度5以上の概念のみを対象にし、コーパス頻度5以上のものが2概念以下の場合、コーパス頻度0もしくは1の概念以外すべてを対象に作業を行った(表2の範囲では「\*」部分の行が調査対象行となる)。こうして選択されたグループが158組、延べ概念数にして555個(同音異義語のリストで一定以上の頻度のものを選択した概念数)となった。

### 3.1.4 概念ごとの格助詞組み合わせ異なりパターンの抽出

上記条件で抽出した概念について、「日本語動詞の結合価」[1]データから概念ごとに格助詞パターンを出力したものが表3である。

表3 (表記+概念ID)グループごとの格助詞パターン

#(見出し+概念ID)グループ	#単語見出し	頻度	表記(語幹)	概念ID	概念説明	格数	格名1	格名2	格名3
1	ア・ウ[会う]	48	会	3ce8ca	会う	1	NI		
1	ア・ウ[会う]	32	会	3ce8ca	会う	2	NI	DE	
1	ア・ウ[会う]	31	会	3ce8ca	会う	1	TO		
1	ア・ウ[会う]	31	会	3ce8ca	会う	2	GA	NI	
1	ア・ウ[会う]	25	会	3ce8ca	会う	2	GA	TO	
1	ア・ウ[会う]	15	会	3ce8ca	会う	3	GA	DE	TO
1	ア・ウ[会う]	12	会	3ce8ca	会う	1	DE		
1	ア・ウ[会う]	11	会	3ce8ca	会う	-1			
1	ア・ウ[会う]	11	会	3ce8ca	会う	1	GA		
1	ア・ウ[会う]	11	会	3ce8ca	会う	2	DE	TO	
1	ア・ウ[会う]	10	会	3ce8ca	会う	3	GA	NI	DE
1	ア・ウ[会う]	7	会	3ce8ca	会う	2	GA	DE	
1	ア・ウ[会う]	1	会	3ce8ca	会う	3	NI	DE	TO
1	ア・ウ[会う]	1	会	3ce8ca	会う	2	NI	TO	
2	ア・ウ[合う]	1	合	0eb828	(二つ以上のものが)一つに合わさる	2	GA	TO	
3	ア・ウ[合う]	25	合	1fa2c9	ぴったり一致する	1	NI		
3	ア・ウ[合う]	16	合	1fa2c9	ぴったり一致する	1	GA		
3	ア・ウ[合う]	11	合	1fa2c9	ぴったり一致する	2	GA	NI	

### 3.2 格助詞組み合わせグループの認定と基本格パターンの設定

#### 3.2.1 概念単位内の格助詞組み合わせのグルーピング

表2で\*のついた概念を対象にして、概念単位ごとに表3の形で出力した格助詞組み合わせパターンどうしを検討し、格助詞の組み合わせが一致していなくても同じパターンから派生していると思われるものと同じグループIDをつける。なお、着目の格助詞組み合わせパターンがすでにグルーピングした複数のグループの省略形候補と思われる

場合は、可能性のある複数のグループIDをつける。

(表4のグループ番号の列で「12」と表示されているものは、グループ1もしくはグループ2の省略形の可能性があることを示す。)

作業過程において、疑問の生じた部分については、元の例文つきデータ(表1相当)に遡って概念ごとに参照しながら作業した。特に表4で示す「x」や「?」のついた部分は元データ参照が必要となった部分である。

### 3.2.2 基本格パターンと省略形の選別

3.2.1で同じグループと判定した中で基本格パターンと思われるものに\*印をつける。

**基本格パターンと省略形の判断にあたって：**

- ・省略形で使われることが一般的と思われても、その省略形の元の形を基本格パターンとする。
- ・着目の基本格パターンが、同じグループ内で、それより組み合わせ数の多い格助詞をとっている他のパターンの省略形と同じになったとしても、基本格パターンの一つとして判断できるものは他の省略形の一致の可能性にかかわらず基本格パターンを設定する。

**その他、作業上で問題があった場合の区分記号の記入について：**

- @：省略形と思われる格の組み合わせのみが出現し、出現したパターンには基本格パターンと思われるものがない場合は、作業者が足りない格を補って、基本格パターンとして追加する。  
作業者が作成したパターンということで\*印の代わりに@印をつける。
- x：あきらかにコーパスの間違い(係り受けの間違いや、動詞の概念同定の間違い)からできたとと思われる格助詞パターン
- ?：係り受け関係などがコーパスレベルで誤って認定された可能性のある格助詞パターン  
この場合は基本的に「日本語動詞の結合価」[1]の元データに遡り、出現した例文にあたり正誤を確認する。

表4 基本格パターン設定結果の表

グループ番号	基本格パターンのサイン	格助詞組み合わせ	頻度	見だし	概念ID	概念
1		_TO	31	ア・ウ[会う]	3ce8ca	会う
1		_GA_TO	25	ア・ウ[会う]	3ce8ca	会う
1	*	_GA_DE_TO	15	ア・ウ[会う]	3ce8ca	会う
1		_DE_TO	11	ア・ウ[会う]	3ce8ca	会う
2		_NI	48	ア・ウ[会う]	3ce8ca	会う
2		_NI_DE	32	ア・ウ[会う]	3ce8ca	会う
2		_GA_NI	31	ア・ウ[会う]	3ce8ca	会う
2	*	_GA_NI_DE	10	ア・ウ[会う]	3ce8ca	会う
12		_DE	12	ア・ウ[会う]	3ce8ca	会う
12		_GA	11	ア・ウ[会う]	3ce8ca	会う
12		_GA_DE	7	ア・ウ[会う]	3ce8ca	会う
?		_NI_DE_TO	1	ア・ウ[会う]	3ce8ca	会う
?		_NI_TO	1	ア・ウ[会う]	3ce8ca	会う
1	*	_GA	16	ア・ウ[会う]	1fa2c9	ぴったり一致する
2	*	_GA_NI	11	ア・ウ[会う]	1fa2c9	ぴったり一致する
2		_GA_NI_DE	1	ア・ウ[会う]	1fa2c9	ぴったり一致する
2		_GA_NI_KR	1	ア・ウ[会う]	1fa2c9	ぴったり一致する
2		_HA_GA_NI	1	ア・ウ[会う]	1fa2c9	ぴったり一致する
2		_NI	25	ア・ウ[会う]	1fa2c9	ぴったり一致する
3	*	_GA_TO	10	ア・ウ[会う]	1fa2c9	ぴったり一致する
123		_DE	1	ア・ウ[会う]	1fa2c9	ぴったり一致する
123		_HA_GA	1	ア・ウ[会う]	1fa2c9	ぴったり一致する
123			2	ア・ウ[会う]	1fa2c9	ぴったり一致する

### 3.3 基本格パターンごとに動詞の概念をまとめる

3.2で作成した表4の中で\*のついた基本格パターンを取り出し、格助詞組み合わせパターンごとに動詞事例を

出力したものが表5である。動詞の単語事例については、単語数に示された数だけ配置されているが、紙面の都合上、各パターンごとに10ずつ表示してある。

表5 基本格パターン別の単語リスト

基本格パターン	単語数	動詞の単語事例
_GA	85	デ・ル[出る],デキ・ル[できる],アル[ある],チガ・ウ[違う],ハッセイ・スル[発生する],ハタラク[動く],ノコル[残る],オワ・ル[終わる],ミエ・ル[見える],タカマ・ル[高まる]
_GA_DE	15	クラ・ス[暮らす],トウジョウ・スル[登場する],オキ・ル[起きる],アツマ・ル[集まる],オワ・ル[終わる],アル[ある],オワ・ル[終わる],オキ・ル[起きる],クラ・ス[暮らす],ウケル[受ける]
_GA_DE_TO	1	ア・ウ[会う]
_GA_HE	35	カワ・ル[変わる],オズレ・ル[訪れる],ツナガ・ル[つながる],キエ・ル[消える],イ・ク[行く],サンカ・スル[参加する],キコク・スル[帰国する],アツマ・ル[集まる],ノリダ・ス[乗り出す],ツク[着く]
_GA_HE_KR	6	デ・ル[出る],モド・ル[戻る],ト・ブ[飛ぶ],トドク[届く],ウツル[移る],オチ・ル[落ちる]
_GA_KR	1	ナガレ・ル[流れる]
_GA_KR_MD	5	デ・ル[出る],モド・ル[戻る],ト・ブ[飛ぶ],ツナガ・ル[つながる],ウツル[移る]
_GA_MD	11	イ・ク[行く],ヒロガ・ル[広がる],カエ・ル[帰る],カヨウ[通う],タツ・スル[達する],セマ・ル[迫る],ス・ム[進む],セマ・ル[迫る],タツ・スル[達する],セマ・ル[迫る]
_GA_NI	111	カワ・ル[変わる],ツナガ・ル[つながる],ナラ・ブ[並ぶ],トモナウ[伴う],ニル[似る],チュウモク・スル[注目する],カワ・ル[変わる],キエ・ル[消える],サンカ・スル[参加する],サンカ・スル[参加する]
_GA_NI_DE	2	ア・ウ[会う],ア・ウ[違う]
_GA_NI_KR	7	デ・ル[出る],モド・ル[戻る],トドク[届く],ウツル[移る],ト・ブ[飛ぶ],オチ・ル[落ちる],モド・ル[戻る]
_GA_NI_TO	2	コタエ・ル[答える],コタエ・ル[答える]
_GA_TO	21	チガ・ウ[違う],ツナガ・ル[つながる],ニル[似る],イ・ウ[いう],イチ・スル[一致する],シュチョウ・スル[主張する],セツゾク・スル[接続する],シュチョウ・スル[主張する],タイオウ・スル[対応する],ア・ウ[会う]
_GA_WO	251	ハッピーウ・スル[発表する],エ・ル[得る],リヨウ・スル[利用する],カイハツ・スル[開発する],キメ・ル[決める],シル[知る],ヒラ・ク[開く],カ・ウ[買う],コエ・ル[超える],デ・ル[出る]

#### 4 作業過程でできた問題点

(1) 元コーパスで概念選択が誤っている事例がある。

サ変動詞では、例1のように自動詞他動詞の両方が存在するものがある。EDR 電子化辞書[2]ではこれらの概念の違いは本来単語辞書レベルでは区別されている。ところがコーパス段階での概念選択(コーパスに出てくる単語の概念がなんであるかを判定し、単語辞書の概念につなぐこと)が違っているものは、本来なら異なる概念なのに自他動詞とも同じ概念に配置されていることがある。自他動詞の場合は、格関係そのものも違ってくる場合が多いので、元々コーパスレベルで概念選択が誤っているものは今回の作業対象外とした。

例1 展開する(他動詞) ~が~を展開する。 展開する(自動詞) ~が展開する。

(2) 着眼の基本格パターンが、ほかの格の組み合わせの省略形であることがある。

例2 ~が(品質・水準)を保つ。 ~が~を(品質・水準)に保つ。

このとき、ガ格とヲ格の組み合わせで出現した場合、の基本格パターンに該当するのか、の省略形なのかは、元の例文つきデータまで、遡らなければならない。

(3) 基本的な格助詞の組み合わせでもコーパスに出現しない場合がある。

コーパスでは、必須格と思われるものでも、省略されてしまう事例が多々ある。そのため省略可能と考え、本来ならあるべき格を省いたまま基本格パターンとみなしてしまうこともある。コーパス依存の作業ではあるが、ある程度、動詞の意味的分類などにより、どのグループだったらどの格助詞が必要という大きな枠組みに対する観点である。例えば、移動の動詞であれば、「始点」になる「から」、「終点」になる「まで、に」などがこのような捉え方である。

(4) 同じ概念でも、始点と終点を示す格をとりうる場合、どちらに焦点があるかによって格助詞のとりかたが変わってくるものなどがある。こういうタイプのものは基本格パターンの設定をどうするかで難しい場合がある。

例3 その女が雑踏に消える。 その女が視野から消える。

これらの二つを「~が~から~に消える」という基本格パターンで扱うか、「~が~に消える」「~が~から消える」と別パターンで扱うか判断に迷うところであるが、の場合はともかくについては、かならずしも終点については言及しなくても文意が成り立つということで、「~から~に」のパターンとは別にパターンを立てる。

文献 [1] 荻野孝野・小林正博・井佐原 均, 「日本語動詞の結合価」, 三省堂(2003.12)

[2] [http://www2.crl.go.jp/kk/e416/EDR/J\\_index.html](http://www2.crl.go.jp/kk/e416/EDR/J_index.html)