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

荻野 孝野^{†,††} 植田 禎子^{††}
小林 正博^{††} 井佐原 均^{†,†††}

係り受け関係のついた大量のコーパスを元にして作成されたデータを対象として、動詞の結合価に関する検討を行った。これは、係り受け関係まで付与された大量データからなるコーパスが存在してはじめて可能となった分析である。動詞の結合価に関する検討は、各動詞の基本的な格パターンに着目して結合価を決定することを中心として検討されてきた。しかし、省略を含め、結合価が実際の言語データでどういう形で出現しているのかについて、全容を示すようなものは報告されていない。ここでは、大量のコーパスデータから作成した結合価データを用い、実際のデータで動詞にかかる格助詞がどういうパターンで出ているのかを調査し、格助詞パターンの出現状況を把握するとともに、それらの格助詞パターンを用いて同音異表記がどの程度判定できるかを検討した。動詞約 12,400 概念 (表記の異なりレベルで約 9,400 単語) から作成した動詞の格助詞組み合わせパターンは、延べパターン数で 37,237 パターン、異なりパターンで 188 パターンとなった。また、同音異表記セットについて、これらのパターンを用い、表記確定を試みたところ、結合価のうち格助詞組み合わせパターンの異なりによる判定でも格助詞パターンの出現頻度などを判定基準に付加することによって約 73% の判定が可能であることがわかった。

キーワード: 日本語, EDR コーパス, 結合価データ, 動詞, 格助詞, 異表記判定

Valence behavior at the level of particle combinations based on corpus data and its contribution to homophone distinction

TAKANO OGINO^{†,††}, YOSHIKO UEDA^{††}, MASAHIRO KOBAYASHI^{††}
and HITOSHI ISAHARA^{†,†††}

The study investigated the behavioral patterns of case-marking particles by using the valency data extracted from a large-scale authentic corpus data. In addition, the study has examined the extend to which such patterns of particle behavior contribute to homophone distinction. 12,400 verb concepts were included in the data, from which combination patterns of case-marking particles were generated. The token number of particle combination patterns was 37,237 in total, 188 being the number of the types. An experimental homophone distinction, in which the orthographic representations of paired homophone words was identified, was conducted by using the generated particle patterns. The results indicated that 73 % of the target homophone pairs were correctly judged at the level of case-marking particles by adding the frequency of particle combination patterns to the judgment criteria.

KeyWords: *Japanese, EDR Corpus, valence data, verbs, case-marking particles, distinction of homophones*

1 はじめに

動詞と動詞にかかる格助詞の関係には、例えば、「{ 子供 } に { 本 } を渡す」「{ 家 } を壊す」という文における「渡す」が「～に～を」, 「壊す」が「～を」と必要とするというように、動詞側から必要とする格助詞に特定の関係が見られ、これを動詞と格助詞の結合価と呼ぶ。英語を日本語に機械翻訳する、あるいは外国人がはじめて日本語を学習する、あるいは文脈理解で省略された格関係が何であるかを前後の文脈から自動的に推定し補うという時に、動詞ごとに必要としている格関係の把握が必要となる。これを本論文では、結合価の部分構成する格助詞パターンの検討として扱う。

著者らは、係り受け関係のついた大量のコーパスを元に作成されたデータを対象として、結合価の部分である格助詞パターンに着目した検討を行った。

従来、大量の自然言語データそのものから、係り受け関係を自動認定し、数量的に結合価の体言部分の意味的傾向を把握する研究などが行われてきた(平岡, 松本 1994; 河原, 黒橋 2002)。しかし、自動的な解析だけでは複雑な係り受け関係が解析できないところがあったり、表層的に係助詞「は、も」で表されるものを意味構造的な観点から格助詞に置き換えて考えなければ、動詞が必要としている格関係の数量的実態を把握しきれないという問題もあった。そこで、大量のコーパスから作成した手作業チェック済みの「日本語動詞の結合価」データ(荻野, 小林, 井佐原 2003)を用いることによって、日本語動詞の結合価パターンの数量的実態を初めて明らかにした。また、得られた結合価データを同音異表記判定に利用することによって、データの有効性について分析した。

以上の詳細について、2章で本研究の背景、3章で調査対象データの概要、結合価データ抽出手順、動詞全体の結合価パターン分布状況など、4章で同音異表記の判別状況などについて述べる。

2 本研究の背景

2.1 言語処理における動詞結合価の利用

自然言語処理において、動詞の結合価は、今まで

- ① 係り受け関係にあいまい性がある場合の判定

† 神戸大学大学院自然科学研究科, Kobe University Graduate School of Science and Technology

†† 日本システムアプリケーション, Japan System Application Co., LTD

††† 独立行政法人情報通信研究機構, National Institute of Information and Communications Technology

- ② 格助詞選択が不適切な文の校正
- ③ 必須の格要素が省略されているときの推測
- ④ 同じ表記の単語に対して複数の訳語が対応しているときの判定
- ⑤ 同じ読みの単語に対して複数の表記が対応しているときの判定

などにおいて利用されてきている。

結合価データを利用した研究やシステム開発は、すでに 1970 年代半ばから行われ、言語処理における一般的な手法として現在でも係り受け関係の複数解の絞り込みや省略された格の推定などに言語データとして活用されている。

例えば、機械翻訳の分野では、訳語選択向上のために作成した「日本語語彙大系」(池原、宮崎、白井、横尾、中岩、小倉、大山、林 1997) の研究があげられる。さらに動詞と名詞の訳語選択における結合価文法利用について、金出地らによって、その効果を定量的に計測した研究(金出地、徳久、村上、池原 2003) などがある。その他、格関係の類似を利用した語義の分類(峯、實成、中山、久保山、谷口、雨宮 1995) や格関係対応を利用した言い換え表現に関する研究(鍛冶、河原、黒橋、佐藤 2003) などもある。他に文解析への利用として、「構文・意味解析と統合した形態素解析に関する研究」(相場、奥村 1995) などがある。結合価を用いた日本語処理に関して、概略をまとめているものとして「結合価構造に基づく日本文解析」(林 1987) などがある。

2.2 本研究で扱う結合価のレベル

石綿敏雄によると、「結合価 (Valenz) ということばが使われたのは、ドイツのテニエルが最初である」とされている(水谷、石綿、荻野、賀来、草薙 1983)。結合価に着目した初期の段階では、「結合価」は「必須成分」「随意成分」を示す格レベルで考えられていて、その後の研究として発展させたヘルビヒとシュンケルによって「結合価と分布(ここで分布とは、従属語の意味特徴を示すことによって、動詞と従属語の選択制限を示すこと)を組み合わせで記述することになった」とある(水谷他 1983)。なお、ここで意味特徴とは、例えば「食べる」にかかる「を」格を伴った体言の意味的制限を意味的グループ「食べ物」で表現するような、類型化した意味情報を指す。

○ 格助詞パターンレベルの結合価の検討

上記の歴史的背景に基づき、ここでは、結合価記述のレベルを以下の 2 段階にわけてとらえる。

- (1) 第 1 段階として、格助詞パターンレベルのもの
- (2) 第 2 段階として、格助詞に導かれる体言レベルの意味的制限まで入れたもの

本論文ではまず上記の (1) の格助詞パターンレベルについて、大量の言語データを対象とし

て、その実態を調査し考察する。次に(1)の検討をさらに発展させた形での(2)のレベルについてもその後調査を行っているが、これは別稿での報告とする(荻野, 吉田, 小林, 井佐原 2005)。

なお、本研究でまず第1段階である格助詞パターンのみに限定した検討を行ったのは、以下のような背景に基づくものである。

結合価に着目した当初は、上記でふれたように、まず格助詞関係のみを扱っているが、その後は、格が担っている体言部分の意味的な制限(選択制限)も含んで扱う形式となり、現在は体言の意味制約まで含めて記述するのが一般的である。しかし、これは解析対象文に新たに出現した体言が、すでに結合価の基準データとして記載された意味的な制限のどれに相当するかが判定できなければ使うことができない。つまり、意味的選択制限で意味情報を自動処理に用いるためには、基本的に出現するすべての体言が、動詞側の結合価の体言部分に付された意味情報と照合可能でなければならない。それには、基本的に結合価の記述のための意味情報を付与した体言辞書が存在しなければならない。従って、未分類の新語が出現した場合や、複合語など解析時に認定される単位と辞書で登録されている単位のずれが生じた場合など、出現した体言に意味情報がつかなかった場合の対応も考えなければならない。

そこで著者らは、結合価の体言部分の情報が未登録で意味マーカを使えない場合でも、あるレベルまでは利用できる格助詞のみの結合価データを大量データから構築することをめざした。

以上の背景に基づき、先にあげた結合価の第1段階である格助詞パターンレベルに限定して、以下の内容で本研究を行った。

① コーパスに出現する格助詞パターンの数量的検討

大量の動詞の実例を分析して、出現する格助詞がいったいどれぐらいの組み合わせパターンに集約できるか、またそれらの格助詞組み合わせパターンを、「が、を、に」など基本的な格助詞の範囲に限定してみたときに、各動詞内の格助詞パターンの出現頻度の分布がどのようになるか。異なり約12,400の動詞概念について係り受け関係を記述したデータから格助詞関係のすべてを抽出し、異なりパターンの数量的検討を行った。

② 格助詞の組み合わせレベルでの異表記判定による評価

①のデータから、同音異表記に該当する事例を抽出し、それらの事例を同じ読みになる動詞グループで分け、格助詞組み合わせの違いだけで、どれぐらい同音異表記の弁別が可能かを検討した。

○ コーパスの出現形に基づく格助詞パターンの検討

動詞の結合価を実際に記述したものとして、石綿の研究(石綿 1975)に基づく結合価データ(水谷他 1983)、さらにIPAL 動詞辞書(情報処理振興事業協会 1987)、NTT グループの構文意味辞書(白井, 横尾, 中岩 1997)などがあげられる。これらはいずれも結合価の利用という立場で、「着目の動詞についての結合価の基本パターンを作成する」という観点で、各動詞の必須

格と思われる格の組み合わせからなる基本的なパターンを作成することが中心に行われてきた。動詞ごとの基本的なパターンを設定することは、利用分野の立場からみて係り受け関係のあいまいな部分の判定や誤った格助詞の校正などには重要な観点である。しかし、同音異表記の判定などにおいては、表現文には必須格相当のものがすべて整った形で出現するとは限らず、必須格とみなされるレベルでも省略される可能性は大きい。したがって、省略される格の状況も含めて、各動詞に格助詞がどう出現するかの傾向をコーパスの現状から把握しておくことも必要と考え、本研究での評価はまずコーパスでの出現状況から行うものとした。

3 動詞の結合価に関する調査

3.1 調査対象のデータ

本調査の対象となったデータは、EDR コーパスから作成された EDR 共起辞書 (日本電子化辞書研究所 1995) を元に、約 12,400 概念の動詞にかかるすべての格関係のデータを抽出して作成された「日本語動詞の結合価」データ (例 2, 附録 A) である

(1) 「日本語動詞の結合価」の元データ：EDR コーパス

例 1 は本稿の検討対象となった「日本語動詞の結合価」データの初期データである EDR コーパスの事例である。EDR コーパスは、形態素情報、構文情報、意味情報を網羅している。本稿の「日本語動詞の結合価」データにかかわる部分は、係り受け関係を付与した構文情報部分である。

例 1 EDR コーパス

```
%%レコード番号%%      (コーパスレコードの識別子)
JC00100520
%%テキスト番号%%      (共起辞書の例文情報に書かれているテキストを限定する番号)
00050003e893
%%出典情報%% (出典)
新聞 870530
%%原文%% (原文)
```

屋内に戻り、お菓子を食べ始めたが、ツカダは遠慮して縁先に立っていた。

```
%%構成要素情報%% (例文の構成要素である形態素の情報)
```

```
1 屋内 オクナイ 名詞 0e757a
2 に ニ 助詞 2621d5
3 戻 モド 動詞 3cfafc
4 り リ 語尾 2621cc
:
```

```
%%形態素情報%% (形態素の表記情報の並び)
```

```
/1:屋内/2:に/3:戻/4:り/5:、/6:お/7:菓子/8:を/9:食/10:始め/11:
:
```

```

%%%構文情報%%% (該当例文の構文構造)
(S (t (M (S (t (S (t (M (S (t (W 1 "屋内"))
                        (W 2 "に"))
                        (t (S (t (W 3 "戻"))
                            (W 4 "り"))))))))
        (W 5 "、"))
    (t (M (S (t (S (W 6 "お")
                (t (W 7 "菓子"))))
        (W 8 "を"))
    (t (S (t (W 9 "食べ"))
:

%%%意味情報%%% (該当例文の意味構造)
[ [main 21:立:0fc282]
  [attribute past state]
  [agent 14:ツカダ:"=Z ツカダという人"]
  [place 19:縁先:0e6a21]
  [condition 16:遠慮:3cec2d]
  [and [ [main 9:食べ:3bc6f0]
        [attribute already begin]
        [object 7:菓子:0e96fa]
%%%管理情報%%%
DATE="95/7/12"

```

(2) 「日本語動詞の結合価」データ

本研究の分析対象データである、「日本語動詞の結合価」データについて、以下に概要を示す。

(2.1) 「日本語動詞の結合価」データの形式

例2は、例1に示すEDRコーパス、そして共起辞書を経て作成された「日本語動詞の結合価」データの部分で、動詞にかかる体言部分を所定の格助詞の列に収めた形式で整理したものである。

例2 結合価データの部分 (詳細は附録A参照)

	A	B	C	D
1	食べる[タベ・ル]			
2	概念ID:[3bc6f0]			
3	概念説明:[食物をとる]			
4	は	が	を	に
5		が(牛(名詞))	→(エサ(名詞))	
6			を(サンドイッチ(名詞))	は(昼食(名詞))
7				
8		は(ヒラメたち(複合語))	を(アジ(名詞))	
9		は(わたし(名詞))		
			を(メシ(名詞))	

(2.2) 「日本語動詞の結合価」データにおける動詞の単位

「日本語動詞の結合価」データに表示されている動詞の単位は、「表記と概念」で限定された見出しが1単位となっている。本稿では、この「日本語動詞の結合価」データを利用したため、「表記と概念」で限定された1レコードを動詞の1単位としている。

(2.3) 「日本語動詞の結合価」データにおける格助詞の扱い

格助詞部分は、左から、「は」、「が」、「を」、「に」、「へ」、「から」、「より」、「まで」、「で」、「と」、「その他」の順に列が定められている(詳細は附録 A 参照)。「は」は他の格助詞に代替できない係助詞が、「その他」には、「だんだんと」などのように様態を示す副詞につく語尾で格表示に該当しないもの、あるいは「3 時に」のように時を示す単語に続く助詞で、述語に束縛される格関係ではない連用修飾語などが配置されている。

「日本語動詞の結合価」データの格助詞部分は、元データとなった EDR コーパスから係り受け関係を抽出したものをベースにして作成されたものであるが、結合価データとしてまとめる段階で、手作業にて実データを解釈し、必要な部分には「態の変換」や「係助詞の実質格への置き換え」などを施し、格助詞を実質的な列に配置している。

- ① 態の変換：受け身・使役・可能表現などは、動詞を基本形にした場合の表層格に変換され、体言は本来入るべき格助詞の列に配置されている。「プレゼントが渡された」は「プレゼントを渡す」という基本形とみなされ、「を列」に「が(プレゼント(名詞))」の形で配置されている
- ② 係助詞の変換：係助詞で具体的な格助詞に置き換えられるものは、基本的に実質的な格の列に配置されている。「私はりんごを食べた」は、①の扱いと同様、「が列」に「は(私(名詞))」の形で配置されている。実質的な格に置き換えられない提題の「は」はそのまま「は」の列に配置されている。係助詞「は」の列に残っている実例をみると、例 3、例 4 のように「切る(=一定量を下回る)」や「合う」など状態を示す述語で、「は」「が」を同時に取り、「は」が提題相当のものなどである。

例 3 M1009 というモデル [は]、ユーザー価格 [が]5 万円を切る。

例 4 弟の言うこと [は] つじつま [が] 合っていない。

以上のような事例を除いて、「日本語動詞の結合価」データでは、基本的に実質的な格に置き換えることができる係助詞は格助詞に置き換えて扱っている。従って、本稿の格助詞パターンも、表層的には係助詞で表現されていても、実質格に置き換えられる部分は実質格の格助詞パターンからなる。

3.2 作業手順

「日本語動詞の結合価」データを基データとして以下の作業を行う。

- ① 動詞ごとに格助詞の組み合わせをすべて抽出する。
格助詞の組み合わせの異なりごとに集計し、動詞ごとの格助詞異なりパターンを作成する。
- ② 対象データ内のすべての動詞から抽出した格助詞異なりパターンを動詞全体について

統合する。

- ③ 統合した格助詞パターンをさらにソートし、動詞全体で格助詞パターンが異なりでいくつ存在するかの情報を得る。
- ④ さらに③の異なりパターン別に単語一覧を作成する。

図 1 格助詞組み合わせパターンの抽出手順

3.3 格助詞パターンの様相

3.3.1 異表記セットの格助詞パターンの抽出

3.2 節の作業手順に従って、「日本語動詞の結合値」データから抽出した格助詞パターンを以下に示す。表 1 は、「かな表記：ア・ウ」に該当するもののすべての表記について、格助詞パターンを抽出して表示したものである。これによると、「ア・ウ」の異表記セットの一部である「会う」と「合う」と「遭う」の格助詞の組み合わせが全く同じパターンになるところと異なるところがある。「～が～に」は「会う/合う/遭う」のいずれにも出現しているが、「～が～と」は「会う/合う」のみである。「会う」と「合う」は表記が違っても、格助詞の組み合わせパターンは非常に似ていることがわかる。「～が～より」は表 1 の範囲では「合う」のみである。

3.3.2 格助詞パターンでまとめた動詞一覧

表 2 に格助詞組み合わせ別に単語を収めた動詞の一部を示す。動詞部分は、読み [表記](出現頻度) で表現してある。概念数とは動詞一覧に記載されている単語の数である。実際は、この概念単位は、先に説明した「表記と概念」で限定される 1 見出し相当である。本来、本研究のパターン弁別は概念単位で行われているので、分類表の単語一覧の部分に概念 ID を付与するの

表 1 「あう」の読みを持つ動詞の格助詞パターン

読み [表記]	格助詞パターン	格助詞パターン頻度	概念識別子	概念説明
ア・ウ [会う]	_DE	12	3ce8ca	会う
ア・ウ [合う]	_DE	1	1fa2c9	ぴったり一致する
ア・ウ [会う]	_DE.TO	11	3ce8ca	会う
ア・ウ [会う]	_GA	11	3ce8ca	会う
ア・ウ [合う]	_GA	26	1fa2c9, 3ce55d, 1fa2cc, 3ce9d3, 1fa2ca	ぴったり一致する, きちんと一致する, 互いに~する, 釣り合いがとれる, 正しいものに合う
ア・ウ [会う]	_GA.DE	7	3ce8ca	会う
ア・ウ [合う]	_GA.DE	1	3ce55d	きちんと一致する
ア・ウ [会う]	_GA.DE.TO	15	3ce8ca	会う
ア・ウ [会う]	_GA.NI	31	3ce8ca	会う
ア・ウ [合う]	_GA.NI	27	1fa2c9, 3ce55d, 3ce9d3	ぴったり一致する, きちんと一致する, 釣り合いがとれる
ア・ウ [遭う]	_GA.NI	16	103c7e	思いがけない人や事件と出会う
ア・ウ [会う]	_GA.NLDE	10	3ce8ca	会う
ア・ウ [合う]	_GA.NLDE	1	1fa2c9	ぴったり一致する
ア・ウ [遭う]	_GA.NLDE	5	103c7e	思いがけない人や事件と出会う
ア・ウ [合う]	_GA.NLKR	1	1fa2c9	ぴったり一致する
ア・ウ [会う]	_GA.TO	25	3ce8ca	会う
ア・ウ [合う]	_GA.TO	13	1fa2c9, 3ce55d, 3ce9d3	ぴったり一致する, きちんと一致する, 釣り合いがとれる
ア・ウ [合う]	_GA.YR	1	3ce9d3	釣り合いがとれる
ア・ウ [合う]	_HA.GA	2	1fa2c9, 3ce55d	ぴったり一致する, きちんと一致する
ア・ウ [合う]	_HA.GA.NI	1	1fa2c9	ぴったり一致する

が正しいが、表示上の都合により、該当の概念を持つ単語の形で表示する。出現頻度は実際にその概念が該当のパターンでコーパスに出現した頻度である。たとえば、他動詞である「～が～を食べる」が「～が食べる」だけの形で出現する場合もある。これは、本来なら対象格を取り他動詞になる「食べる」が、対象格を取らない一般的な自動詞のパターン「～が」と同じパターンになってしまうということである。動詞ごとに必須の格が決定されるとすると、本来は必須格でみた格助詞パターンがいくつ存在するかという立場での集計も重要であるが、ここではあくまでもコーパスに出現する形での異なりパターン一覧表である。これは、コーパスに出現する形を対象にどう解析するかということを検討する素材となる。なお、動詞一覧の単語の後ろの()に、該当のパターンで出現したコーパス中での頻度を示す。

附録 B に、異なりパターンレベル上位頻度からの格助詞パターンと該当する単語の一部を例示する。

3.3.3 格助詞組み合わせの異なりパターン

(1) 格助詞パターンの異なり数

「日本語動詞の結合価」に出現した動詞概念の事例約 157,000 について、格助詞の組み合わ

表 2 格助詞パターン別動詞一覧 (他のパターンは付録 B 参照)

格助詞パターン	概念数	事例頻度合計	動詞一覧
_GA	4,016	22,141	ア・ウ [会う] (11) ア・ウ [合う] (16) ア・ウ [合う] (1) ア・ウ [合う] (2) ア・ウ [合う] (5) ア・ウ [合う] (2) ア・ク [開く] (5) ア・ク [開く] (1) ア・ク [空く] (2) ア・ク [空く] (1) ア・ク [空く] (2) ア・ク [空く] (1) ア・ク [空く] (2) ア・ム [編む] (1) ア・ル [ある] (1) ア・ル [ある] (14) ア・ル [ある] (332) アイ・スル [愛する] (3) アイ・スル [愛する] (1) アイコウ・スル [愛好する] (1) アイコトナ・スル [相異なる] (1) アイサツ・スル [挨拶する] (2) アイサツ・スル [挨拶する] (1) アイズ・スル [合図する] (2) アイツ・ク [相次ぐ] (86) アイハン・スル [相反する] (1) アオ・ル [あおる] (1) アオザメ・ル [青ざめる] (1) アカスケ・ル [あか抜ける] (1) アガ・ル [挙がる] (3) アガ・ル [上がる] (1) アガ・ル [上がる] (6) アガ・ル [上がる] (1) アガ・ル [上がる] (105) アガリコ・ム [上がり込む] (1) アガリツツケ・ル [上がり続ける] (2) アガリハジメ・ル [上がり始める] (2) アキラメ・ル [諦める] (3) アクシユ・スル [握手する] (2) アクセス・スル [アクセスする] (1) アグネ・ル [あぐねる] (1) アケ・ル [明ける] (13) アケ・ル [明ける] (7) アケ・ル [明ける] (4) アゲ・ル [挙げる] (1) アジワエ・ル [味わえる] (1) アセ・ル [あせる] (2) アセ・ル [焦る] (1) アソ・ブ [遊ぶ] (1) アソ・ブ [遊ぶ] (14):
_GA-WO-NI	1,003	3,222	アイヨウ・スル [愛用する] (1) アオ・グ [仰ぐ] (1) アカ・ス [明かす] (1) アケ・ル [開ける] (1) アケワタ・ス [明け渡す] (1) アゲ・ル [挙げる] (2) アゲ・ル [上げる] (1) アゲ・ル [上げる] (2) アゲ・ル [上げる] (1) アズカ・ル [預かる] (1) アズケ・ル [預ける] (8) アズケ・ル [預ける] (1) アタエ・ル [与える] (6) アタエ・ル [与える] (25) アタエ・ル [与える] (20) アツカ・ウ [扱う] (1) アツマ・ル [集まる] (1) アツメ・ル [集める] (6) アツメ・ル [集める] (6) アテ・ル [当てる] (1) アテ・ル [当てる] (2) アテ・ル [当てる] (1) アビ・ル [浴びる] (1) アビ・ル [浴びる] (2) アビ・ル [浴びる] (1) アビセ・ル [浴びせる] (1) アピール・スル [アピールする] (1) アユ・ム [歩む] (1) アララゲ・ル [荒らげる] (1) アラワ・ス [著す] (1) アル・ク [歩く] (1) アワセ・ル [合わせる] (6) アワセ・ル [合わせる] (1) イ・ウ [いう] (1):

表 3 格助詞パターンの延べパターン数と異なりパターン数

調査した概念数	12,416 個
調査対象に含まれた事例数	156,959 個
延べパターン数	37,237 個
全体の異なりパターン数	188 個

せを抽出し、異なりパターンを出した。約 37,000 の延べパターンは、異なりパターン 188 パターンとなった。

(2) 格助詞パターンの頻度

表 4 に 188 異なりパターンのうち、出現概念の多い上位の格助詞パターンを表示する。

(3) 一つの動詞がとる格助詞パターンの数

表 4 出現頻度の多い格助詞パターン

格助詞パターン	事例頻度合計	事例の割合	概念異なり数	異なり概念の割合
_WO	32414	21 %	4,502	12 %
_GA_WO	23868	15 %	3,820	10 %
_GA	22141	14 %	4,016	11 %
_GA_NI	13,352	9 %	2,270	6 %
_NI	9031	6 %	1,804	5 %
_WO_NI	8,624	5 %	1,694	5 %
_WO_DE	7,965	5 %	2,058	6 %
_GA_DE	5,566	4 %	1,739	5 %
NIL	4915	3 %	2086	6 %
_GA_WO_NI	3,222	2 %	1,003	3 %
_GA_WO_DE	3,173	2 %	1,223	3 %
_DE	3,001	2 %	1,310	4 %
_TO	2,079	1 %	366	1 %
_WO_KR	1,501	1 %	611	2 %
_GA_TO	1,485	1 %	413	1 %
_GA_KR	1,287	1 %	477	1 %
_WO_TO	1,247	1 %	219	1 %
_GA_NI_DE	939	1 %	492	1 %

1つの動詞概念がとる格助詞パターンは、1概念で平均して3.3パターン(表3の調査した概念数に対する延べパターン数の割合)である。最大の異なりパターンを持つものは、「運ぶ」でパターン数42個である(表5)。出現した格助詞パターンが多い動詞は、コーパス中での出現頻度が高いものや、意味的に「移動を表す」範疇に入る動詞概念が多い。移動を表す動詞で格助詞パターンの種類が豊富なのは、行き先をあらわす格助詞として、同じ役割を表現できる格助詞が「に、へ、まで」などと複数個あるために、他の種類の格助詞と合わせて、格助詞パターンが組み合わせ的に増えるためと思われる。

3.3.4 格助詞パターンからみた動詞の大分類

表7は、「格名1」から「格名6」までをそれぞれ「は、が、を、に、へ、から、より、まで、で、と」の順にソートした上で、以下に示す基準に従って「は、が、で」など、どの動詞グループにも出現可能な係助詞、格助詞を除き、結合価パターンを特徴づける格助詞と思われる「を、に、から、より、へ、まで、と」を優先して、その順に大分類(第1分類)、小分類(第2分類)の種別を入れたものである。これは、動詞を特徴的に分類するにあたって、重要度の高いと思われる格助詞の優先順をつけて動詞を分類し、それぞれのグループの比率を把握するために行ったものである。こうして格助詞組み合わせをおおまかにとらえたグループは、「対象格を取るグループ」、「始点終点を取るグループ」というような動詞のグループにも関連付けられる。

(1) 格助詞の置き換え

表 5 概念別パターン数でみた異なりパターン数上位の概念

# 単語見出し	概念識別子	概念説明	同じ概念内のパターン数	頻度
ハコ・ブ [運ぶ]	1e85e6	物を持ったり載せたりして、他の場所に移す	42	43
デ・ル [出る]	1e85b5	(ある場所の) 内から外へ行く	36	70
スス・ム [進む]	3ce9f3	前へ進む	31	24
デキ・ル [できる]	1f1d34	物事を行うことができる	31	533
ト・ブ [飛ぶ]	105333	飛行する	31	34
ハイ・ル [入る]	3ce9c4	ある場所の中に入る	31	157
モチイ・ル [用いる]	3cf750	役立たせ、うまく運用、採用する	31	586
ウ・ル [売る]	3cfd6c	品物を売りさばく	30	87
オク・ル [送る]	3cff8b	人を、何かの必要のために目的の所へ行かせる	29	26
ハジメ・ル [始める]	1fb02a	(これまで経験のない動作や物事を) 開始する	29	258
キタイ・スル [期待する]	3ce69b	物事の実現を待ち望む	28	129
ススメ・ル [進める]	3ceaeb	物事を進展させる	28	69
シメ・ス [示す]	3c5a5e	はっきりとわかるように外に表す	27	414
ダ・ス [出す]	3cf7b7	物をさしだす	27	145
デ・ル [出る]	0effae	姿を現わすこと	27	645
アル・ク [歩く]	3cefe6	足を使って歩く	26	88
イドウ・スル [移動する]	3ce645	ある場所から他の場所へ動く	26	23
オコナ・ウ [行なう]	1e84d3	(物事を) 実行する	26	415
カ・ウ [買う]	1e84f1	(品物や権利などを) 代金を払って手に入れる	26	146
ノベル [述べる]	3ce7d8	考えていることを、口に出して言う	26	49
ユシュツ・スル [輸出する]	10cc3d	外国へ向けて産物・生産技術などを送り出す	26	16

①着目の格助詞組み合わせにおいて、格助詞がないものおよび「は、で」だけのものは、特徴的な格助詞がないものとして「NASI」として扱う。

②着目の格助詞組み合わせにおいて、「は、が、で」のいずれかあるいはそれらが複数出現している場合で、「が、で」以外の格助詞が存在するなら、「は、が、で」は抽出対象に含めない。

これは、「は、が、で」は基本的にどの動詞グループにも出現する可能性があるものとして、他の格助詞があるかぎり、「は、が、で」は格助詞組み合わせの中に含めないとするものである。他の格助詞がある場合に「は、が」を除いて格助詞パターンを考えるというのは、日本語表現において、「～が～を動かす」というような対象行為や「～が立ち上がる」というような自発行為の動詞において主体を示す「が」は省略される可能性が高いという想定に基づくものである。「で」は、おおまかにとらえると、場所を表す「で」、手段や道具を表す「で」がある。前者の場所を表す「で」は特に動詞に限定されずに使える任意成分に近い格であり、後者の手段や道具を表

す「で」は動詞に限定される動詞固有の格になる可能性が高い。本来これらの格は区別して記載すべきものと思われるが、元データとした EDR コーパスにおいて格助詞は表層の文字列で認識されているだけで、元のデータにおける意味的機能のレベルまで細分されていないので、「で」を任意格に近いものとし、おおまかな動詞分類をするために、考慮する格から外して検討した。

③「から」と「より」のように機能的に交替可能な格助詞は、より基本的とする方の格助詞に変換する。これは似たような機能を持つ格助詞を代表的な格助詞に置き換え、パターンを単純にするためである。著者らの調査は実際のコーパスから抽出したもので、交替可能な格関係は表層的には別パターンとして抽出されているが、IPAL の「基本動詞辞書」(情報処理振興事業協会 1987) ではこれらの格交替可能な部分を以下のように表現している。「日本語語彙大系」(池原他 1997) でも同様の扱いである。本研究でも、大まかな分類レベルでは交代可能な格として一つに置き換えて集計した。

・格形式の交替

格形式が常に交替できるものについては、「二/へ」などのように併記する。

(37) a 彼が夜道を駅から家に急ぐ。

b. 彼が夜道を駅から家へ急ぐ。

-->N1 ガ N2 ヲ N3 カラ N4 二/へ

引用 (情報処理振興事業協会 p.27)

本稿では、(37)a,b のケースは、コーパスから抽出したパターン (付録 B のレベル) では別パターンとして記述してあるが、以下の (2) 大分類での格助詞パターンの出現状況に示す格助詞パターンを大分類で分ける時には、表 6 に該当するものを置き換えて同じものとして扱った。

表 6 格助詞の置き換えを行うもの

元の格助詞	代表的格助詞に置き換え
YR(より)	KR(から)
HE(へ)	NI(に)

(2) 大分類での格助詞パターンの出現状況

(1) 格助詞の置き換えの①から③の手順で、出現した各格助詞パターンの格助詞の一部を考慮外にしたり、置き換えたりして、そのパターンの中で優先度の高い二つの格助詞部分に着目して、第 1 分類 (大分類)、第 2 分類 (小分類) とし、それぞれの列を下記の優先順にて分類した。

1 を、2 に (へ)、3 から (より)、4 と、5 まで

上記基準に従って格助詞を置き換えたものは表 7 の通りである。

表 7 格助詞パターンからの動詞分類のための格助詞置き換え表 (全体は附録 C 参照)

格名 1	格名 2	格名 3	格名 4	格名 5	格名 6	頻度合計	第 1 分類	第 2 分類
GA	WO	KR				606	WO	KR
GA	WO	KR	MD			11	WO	KR
GA	WO	KR	MD	DE		2	WO	KR
GA	WO	KR	MD	DE	TO	1	WO	KR
GA	WO	KR	DE			51	WO	KR
GA	WO	KR	DE	TO		1	WO	KR
GA	WO	KR	TO			4	WO	KR
GA	WO	YR				22	WO	KR
GA	WO	YR	MD			1	WO	KR
GA	WO	YR	DE			2	WO	KR
HA	GA	WO	KR			2	WO	KR
HA	GA	WO	KR	MD	DE	1	WO	KR
HA	GA	WO	KR	DE		2	WO	KR
HA	GA	WO	YR			1	WO	KR
HA	WO	KR				32	WO	KR
HA	WO	KR	MD	DE		1	WO	KR
HA	WO	KR	DE			1	WO	KR
HA	WO	YR				3	WO	KR
WO	KR	DE				147	WO	KR

表 8 は表 7 で置き換えた第 1 分類, 第 2 分類の異なりパターンごとに頻度を集計したものである。

表 8 格助詞パターン第 1, 第 2 分類レベルの頻度

格数	第 1 分類	第 2 分類	第 2 分類頻度	第 1 分類頻度
1	WO		68,620	
2	WO	NI	13,514	
2	WO	KR	2,470	
2	WO	TO	1,628	
2	WO	MD	236	86,468
1	GA		29,043	
2	NI		25,403	
3	NI	KR	837	
3	NI	TO	158	
3	NI	MD	73	55,514
0	NASI		8,027	8,027
1	TO		4,169	4,169
2	KR		2,270	
3	KR	MD	58	
2	KR	TO	32	2,360
2	MD		415	
2	MD	TO	6	421
	合計		156,959	156,959

図 2 格助詞による動詞分類 (第 2 格助詞まで)

図 3 格助詞による動詞分類 (第 1 格助詞まで)

図2は、表8にまとめた第2分類までの組み合わせで格助詞パターンの割合を図示したもの、図3は第1分類レベルでの割合を表現したものである。

図3の第1格助詞レベルの分類でみると、「物を買う」のように「対象格」を取るものが動詞の約55%、その他は、「雨が降る」のように「現象」を示す動詞や、「歩く、走る」のように対象格を必要としない「が」格だけの動詞が19%、「右に曲がる」のように格助詞の中に「を」を含まず「に」を含む動詞が17%で、以上「が、を、に」のいずれかを第1格にする動詞で全体の約91%になることなどがわかる。なお、「ボタンを右に動かす」の表現には「を、に」が含まれるが、図3は第1格で分類して図示したもので、「を、に」は「を」のグループとして数える。(すべての異なり格助詞パターンおよび第1分類、第2分類でまとめた数の詳細については附録Cを参照のこと。)

3.3.5 動詞内での格助詞パターンのグループ化と基本格の設定

動詞の格助詞パターンは、いつも必要な格がすべて表現された動詞の基本格の形で出現するとは限らない。表現の状態によって、基本的な格助詞パターンの一部が省略されたり、どの動詞にもつきうような格がついている場合がある。

3.3.3節に述べた調査は、基本格パターンと省略格パターンの関係はつけずに、格助詞パターンの出現状況のみでまとめたものである。実際に格助詞パターンを利用するにあたっては、異なる格助詞パターンの基本格パターンと省略格パターンの関連を把握し、それぞれの格助詞パターンが同じグループかどうか、また同じグループとみなされた中でどれが代表的な格助詞パターンであるかの検討が必要となる。

そこで、EDR 単語辞書から抽出した同音異表記および同表記異義語に該当するデータで、動詞の結合価データに一定頻度以上で出現した同音異表記および同表記異義語のセット158組(事例にして5,584事例、概念数で555概念)について、格助詞パターンのグルーピングを試みた。同じパターンから発生したケースと思われるものを1グループとしてグループ分けするとともに、そのグループの中で基本となる格助詞パターンを選定、これを基本格パターンとしたものである。以下、「あう」を事例として、格助詞パターンのグループと基本格の設定について説明する。

(1) 同じ概念内の格助詞パターンと基本格パターンの状況

表9におけるグループ番号の数字は、各グループ番号を示している。同じグループ番号がついている範囲で基本格サインの列で「*」がついているものを基本格パターン、無印のものを同じグループ内の派生形パターンと見なす。表9の事例で「～で～と会う、～が～と会う、～と会う」は、基本格パターン「～が～と 会う」の派生形パターンと見なし、これらはすべて同じ基本格パターンから発生したものとして扱うことを示す。

表 9 同じ概念内での格助詞パターンと基本格パターンの状況

グループ番号	基本格サイン	格助詞パターン	頻度	同概念内割合	読み [見出し]	概念	概念説明
1		_TO	31	12.6%	ア・ウ [会う]	3ce8ca	会う
1	*	_GA_TO	25	10.2%	ア・ウ [会う]	3ce8ca	会う
1		_GA_DE_TO	15	6.1%	ア・ウ [会う]	3ce8ca	会う
1		_DE_TO	11	4.5%	ア・ウ [会う]	3ce8ca	会う
2		_NI	48	19.5%	ア・ウ [会う]	3ce8ca	会う
2		_NI_DE	32	13%	ア・ウ [会う]	3ce8ca	会う
2	*	_GA_NI	31	12.6%	ア・ウ [会う]	3ce8ca	会う
2		_GA_NI_DE	10	4.1%	ア・ウ [会う]	3ce8ca	会う
12		_DE	12	4.9%	ア・ウ [会う]	3ce8ca	会う
12		_GA	11	4.5%	ア・ウ [会う]	3ce8ca	会う
12		_GA_DE	7	2.8%	ア・ウ [会う]	3ce8ca	会う
12		nil	11	4.5%	ア・ウ [会う]	3ce8ca	会う

図 4 「会う」の格助詞パターン

他の概念については、部分的であるが、附録 D に示す。表 9 のグループ番号列で 12 と表示されているものは、グループ番号 1 にもグループ番号 2 にも属することを示す。

表 9 のグループ番号 12 と 1,2 の関係は、図示すると図 4 や図 5 のようになる。重なり合う部分の格助詞パターンについては、グループ 1、グループ 2 両方の省略格パターンである可能性があることを示す。

図 4 は、ひとつの表記における格助詞パターンの重なりであるが、同じ読みで異なる表記があるセットについては、さらに多様性を含むことになる。

図 4 に示す「会う」のグループ 1 と図 5 に示す「合う」のグループ 2 は、「GA_TO」を共有しているので、表記変換上、「会う」と「合う」の二つの可能性がある。同様に、「会う」のグループ 2 と「合う」のグループ 1 と「遭う」のグループ 1 が「GA_NI」を共有して、表記変換上「会う、合う、遭う」の三つの可能性があることがわかる。

(2) 格助詞パターンに関する考察

図 5 「合う」の格助詞パターン

図 6 「遭う」の格助詞パターン

以上、格助詞パターンの異なりやパターンの重なり関係を調査した結果、動詞の概念が持つ格助詞パターンが複数あるとき、その組み合わせには以下のようなタイプが含まれている。

① 基本的なパターンと派生のパターンで多様性が出ている場合

表 9 において、グループ番号と基本格サインの項目で、同じグループ番号内の関係を示す。基本格パターンは「*」で、省略されたパターンあるいは任意格相当の格の付加できたパターンの関係は無印で区別する。たとえば、例 5.1 が基本格パターンになるもの、例 5.2 が任意格が付加された格助詞パターンということになる。

例 5.1 連盟会長が通産大臣に会う。

例 5.2 委員長が、某所で政治家に会った。

② 同じ意味が、異なる格助詞パターンで表現可能な場合

同じ意味的役割を担う置き換え可能な格助詞があるとき、同じ意味が、異なる格助詞の組み合わせで表現されている場合がある。名詞に後続する助詞が違っていても全体の文意が変わらないケースである。「相手」を示す「に、と」、始点を示す「から、より」、終点を示す「まで、に、へ」などのグループでこの傾向がみられる。

◇ (場所終点){に, へ, まで}行く

◇ (対象){に, を}注目する

これらのパターンに対し, 意味的役割を表現する深層格 (国立国語研究所 1997) を用いて表現してみると格助詞が違っていても同じ意味的關係になっていることがわかる.

例 6.1 親 (agent) が僕の彼女 (goal) に会う

例 6.2 親 (agent) が僕の彼女 (goal) と会う

③ 意味も異なって格助詞パターンも異なる場合

格助詞の一部が2項から3項の組み合わせに展開され, 全く別の格助詞パターンになり, 概念関係も異なるというタイプである.

例 7.1 虫 (object) が発生する。(「虫」が自ら変化する現象の主体)

例 7.2 植物 (causal potency) が光合成で酸素を発生する。

(「植物」が現象の原因となる無意志主体)

4 格助詞パターンを用いた同音異表記の判定

3章では, 約 157,000 の事例から抽出した動詞の格助詞パターンにかかわる,

- ① 総パターン数および異なりパターン数など数量的概要
- ② 格助詞パターンに該当する動詞
- ③ 概念内の格助詞パターンのグループ化と基本格の関係

について述べた.

本章では, これらの格助詞組み合わせ情報だけで, 言語処理などにどの程度使えるかの検討を行う. ここでは「同音異表記の判定」の利用を対象に, その可能性を検討する.

4.1 同音異表記判定のためのデータ抽出

(1) 検討対象とする動詞概念候補の抽出 (第1段階)

3.3.5節で抽出した158組は, EDR 単語辞書で, ①同じ読みで複数の表記を持つもの, ②表記の違いの有無に関係なく, 同じ読みで複数の概念を持っているものからなる. ここではそのうち同音異表記 (①同じ読みで複数の表記が存在する動詞概念のグループ) に該当する動詞概念のグループのみを対象とする (表 10).

(2) 検討対象とする動詞概念の絞り込み (第2段階)

(1)で抽出されたデータは, 「日本語動詞の結合価」データの出現頻度を基に以下の基準で検討対象とする範囲を絞り込んだものである.

- ① 同じ読みからなるグループ内で, コーパス出現頻度5以上の動詞概念が3つ以上あるときは5以上の動詞概念のみを対象にする.

- ② 同じ読みからなるグループ内で、コーパス頻度 5 以上の動詞概念が 2 つ以下しかない場合は、頻度 5 以下にも範囲を広げ、コーパス頻度 0 もしくは頻度 1 の動詞概念を除いてすべてを検討対象とする。

上記の選択手順に従って、一定以上の頻度のある動詞概念を選択したところ、表 10 で ○ をつけたところが、本研究での検討対象となるデータで、41 組の同音異表記セットとなった。表 10 の各グループ内は、コーパスに出現した事例の多い動詞概念 (頻度) の順に並んでいる。

表 10 同じ読みで複数の表記が存在するグループ

グループ	作業対象の選択	頻度 (例文数)	読み	表記	語幹	概念 I D	概念説明
1	○	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
		0	ア・ウ	合う	合	1fa2cb	引き合う
		0	ア・ウ	合う	合	1fa2c7	(二つ以上のものが) 一つに集まる
		0	ア・ウ	合う	合	1fa2cd	～して一緒になる
2	○	149	アガ・ル	上がる	上が ^ス	3cf546	前からあるものに別のものを合わせて程度を増す
	○	9	アガ・ル	上がる	上が ^ス	0ee91b	終わりになる
	○	3	アガ・ル	挙がる	挙が ^ス	0e2baa	(犯人が) つかまる
		2	アガ・ル	上がる	上が ^ス	0e2bb5	(部屋に) 上がる
		1	アガ・ル	挙がる	挙が ^ス	0e2bac	(体の一部分が) 上へ動く
		1	アガ・ル	上がる	上が ^ス	0e2bc2	(名声が) 上がる
		1	アガ・ル	上がる	上が ^ス	1fafa8	水から上がる
		1	アガ・ル	上がる	上が ^ス	3c2c8e	緊張して神経質になる
		0	アガ・ル	挙がる	挙が ^ス	0e2bab	(証拠が) 見つかる
		0	アガ・ル	上る	上	3c2c8c	望ましい成果を得る
		0	アガ・ル	上る	上	0e2bc7	検挙される
		0	アガ・ル	上る	上	3c2c8f	一定の費用でまかなえる
		0	アガ・ル	上る	上	3ce776	物事が進行して、ある状態になる
		0	アガ・ル	上がる	上が ^ス	0e560c	飲食する
		0	アガ・ル	揚がる	揚が ^ス	3ce6c7	上昇する
		0	アガ・ル	揚がる	揚が ^ス	3d0464	高い所へ上がる
		0	アガ・ル	揚がる	揚が ^ス	0e2bb6	(遊郭に) 上がる
		0	アガ・ル	揚がる	揚が ^ス	3c2c87	大きな声が上がる

(3) 格助詞パターンの抽出

表 10 の「作業対象の選択」の列で ○ のついた「表記と概念」で特定されたデータについて、

「日本語動詞の結合価」データからまとめた格助詞パターンを抽出する．このようにして，同じ表記内で同じ格助詞パターンになるものをまとめた結果，延べパターンにして1,057パターン，異なりパターンにして112パターンとなった．異表記セットごとの格助詞パターン数の内訳は，表14の各セットに対応する「グループ内格助詞パターン異なり数」および「グループ内格助詞パターン総数」の項目で参照できる．

4.2 格助詞パターンの異なりを用いた異表記判定

4.2.1 格助詞パターンの重なり状況の調査

同じ読みのセットを1グループとした場合，

- ① 格助詞の組み合わせが同じもの
- ② 格助詞の組み合わせは異なるが，他の格助詞組み合わせの省略されたパターンである可能性のあるもの

というように，パターンの重複は，異表記単語間で，単に同じ格助詞パターンがあるかないかだけでなく，省略などによる包含関係も含めてチェックすることが必要である．

事例で見ると以下のようなものである．

- ① 「異表記セット」が同じ格助詞パターンになるケース
これは，同じ格助詞パターンで複数の表記になる可能性があるので，格助詞パターンレベルでは判定不可になるケースである．
- ② 着目の格助詞パターンでは表記一つだが，その格助詞パターンを包含する，より広い範囲の格助詞パターンに「異表記」が存在するケース

表 11 同じ格助詞パターンを持つ異表記セット

表記判定	格助詞パターン	格助詞パターン頻度	読み [表記]
×	._GA_NLDE	10	ア・ウ [会う]
×	._GA_NLDE	1	ア・ウ [合う]
×	._GA_NLDE	5	ア・ウ [遭う]

これは，単独の異なりパターンと表記の関係からは一意に表記が決定するようにみえるが，単に省略格パターンの可能性もあるので，同じグループ内の基本形となる格助詞パターンの重複チェックが必要となるケースである．表12に示す「○ → ×」がその部分で，コーパスの出現状態だけでみた場合は判定可 ○，省略格として出現する可能性も導入した場合には判定不可 ×となることを示す．

「～とあう」で，「と」を取る格助詞パターンは表12に示す「会う」の出現頻度31例のみである．しかし，3.3.5節で基本格パターンと省略格パターンの関係について述べたように，「～と

表 12 省略格パターンの可能性のある場合の判定

×	_GA_TO	25	ア・ウ [会う]
×	_GA_TO	12	ア・ウ [合う]
○ → ×	_TO	31	ア・ウ [会う]

あう」が「～が～と合う」あるいは「～が～と会う」いずれかの省略格パターンになる可能性もあるので、第2段階の判断として、このレベルの格助詞組み合わせまで考慮すると「判定不可 ×」となる。

このように、より小さい格助詞パターンのレベルで表記が一つしか存在しなくて判別可能にみえても、そのパターンを包含する格助詞組み合わせレベルで複数表記ある場合は、より少ない格助詞組み合わせレベルでも「判定不可 ×」とする。

4.2.2 格助詞パターンの違いによる異表記判定

表13の「表記判定」で○の部分は表記の違いに対応して格助詞パターンの違いが出るところ、×の部分は表記の違いによる格助詞パターンの違いが出ていないところである。

表13に示した「あう」の頻度は、399事例、そのうち判別可能な組み合わせは31事例であった。以上のようにして、4.1節の抽出基準で選定された同音異表記に該当する41組の異表記単語のすべてについて格助詞の組み合わせと表記の違いの関係を調査し、これらの範囲で表記による格助詞の組み合わせパターンの違いが出た結果を表示したものが表14である。

判定率上位のものは、「避ける、裂ける」「減る、経る」「煮る、似る」のように、「を格を取る対象行為(避ける、経る、煮る)」と「対象格を取らない現象(裂ける、減る、似る)」の組み合わせなど、動詞の意味的グループからみて、格助詞パターンでの区別が可能な語群のものであることがわかる。

4.2.3 テストデータによる評価

(1) テストデータの抽出と形式化

4.2.2節で日本語動詞の結合価データより作成した同音異表記の判定表(表14)を基準データとして、他のコーパスの該当部分を抽出し、判定を行った。評価データは、京大コーパス(Version3.0, 約40,000文)から抽出した、表14に示した同音異表記41組の読みに該当する単語を含む4,024件のデータである。そのデータ事例として「あう」を格助詞の組み合わせで形式化したものを表15に示す。

(2) 判定データとの照合

京大データから抽出した41組の同音異表記セットについて「日本語動詞の結合価」データから作成した同音異表記セットの判定表(表13)と照合して表記を決定し、元々の表記と一致不

表 13 「あう」に関する格助詞パターンによる判定

同じ読みのグループid	同じ読みに該当するパターン総数	異なり格助詞パターングループid	格パターンにおける異なり表記数	表記判定	同じ読み中の異なり表記数	格助詞パターン	格助詞パターン頻度	読み [表記]
1	29	1	2	×	3	_DE	12	ア・ウ [会う]
1	29	1	2	×	3	_DE	1	ア・ウ [合う]
1	29	2	1	○	3	_DE.TO	11	ア・ウ [会う]
1	29	3	2	×	3	_GA	26	ア・ウ [合う]
1	29	3	2	×	3	_GA	11	ア・ウ [会う]
1	29	4	2	×	3	_GA.DE	7	ア・ウ [会う]
1	29	4	2	×	3	_GA.DE	1	ア・ウ [合う]
1	29	5	1	○	3	_GA.DE.TO	15	ア・ウ [会う]
1	29	6	3	×	3	_GA.NI	16	ア・ウ [遭う]
1	29	6	3	×	3	_GA.NI	27	ア・ウ [合う]
1	29	6	3	×	3	_GA.NI	31	ア・ウ [会う]
1	29	7	3	×	3	_GA.NLDE	5	ア・ウ [遭う]
1	29	7	3	×	3	_GA.NLDE	10	ア・ウ [会う]
1	29	7	3	×	3	_GA.NLDE	1	ア・ウ [合う]
1	29	8	1	○	3	_GA.NLKR	1	ア・ウ [合う]
1	29	9	2	×	3	_GA.TO	25	ア・ウ [会う]
1	29	9	2	×	3	_GA.TO	13	ア・ウ [合う]
1	29	10	1	○	3	_GA.YR	1	ア・ウ [合う]
1	29	11	1	○	3	_HA_GA	2	ア・ウ [合う]
1	29	12	1	○	3	_HA_GA.NI	1	ア・ウ [合う]
1	29	13	3	×	3	_NI	38	ア・ウ [合う]
1	29	13	3	×	3	_NI	8	ア・ウ [遭う]
1	29	13	3	×	3	_NI	48	ア・ウ [会う]
1	29	15	2	×	3	_NLDE	7	ア・ウ [遭う]
1	29	14	2	×	3	_NLDE	32	ア・ウ [会う]
1	29	15	1	○ → ×	3	_TO	31	ア・ウ [会う]
1	29	16	2	×	3	nil	7	ア・ウ [合う]
1	29	16	2	×	3	nil	11	ア・ウ [会う]

一致の度合いを記入したものの一部(「あう」に該当する部分)が、表 16 である。これをすべての組について行い、パターンの一貫度を出したものが表 17 である。

照合は、以下のような手順で行った。

- ① 格助詞パターンを抽出し(表 15)、判定表(表 13)の中で、同じ読みの単位で一致するパターンを探す。一致したパターンで表記が一つしかなければ、第 1 段階で表記が確定したとする。
- ② 第 1 段階で複数の候補が出たものは、判定表で格助詞パターンの頻度をみて、頻度の多い方の表記を選択する。

評価は、表 16 で × のついたものは、判定表から確定できる表記が二つ以上で決定できない、あるいは決定した表記が正解表記に合っていなかったものである。無印のものは表記が一つに確定され、かつ正解の表記と一致したものである。なお、判定表に該当するパターンがない場

表 14 格助詞パターンによる異表記判定の状況

	読み	表記	グループ 内格助詞 パターン 異なり数	グループ 内表記数	グループ 内格助詞 パターン 総数	判定 ○ のついた パターン に該当す る頻度①	同じ読み に該当す るコーパ ス出現頻 度②	頻度でみ た正解率 (①/②)
1	ア・ウ	会, 合, 遭	16	3	29	31	399	7.8%
2	アガ・ル	上が, 拳が	15	2	16	48	162	29.6%
3	アラワ・ス	表わ, 著	11	2	13	85	265	32.1%
4	イ・ク	逝, 行	26	2	31	141	208	67.8%
5	ウツ・ル	映, 写, 移	27	3	36	78	280	27.9%
6	オク・ル	送, 贈	43	2	64	36	476	7.6%
7	カ・ウ	買, 飼	28	2	40	45	572	7.9%
8	カ・ク	書, 欠	24	2	34	16	284	5.6%
9	カエ・ル	変え, 帰, 買え 替え, 代え	46	5	72	57	515	11.1%
10	カンセイ・スル	完成, 管制	13	2	15	139	192	72.4%
11	キ・ル	着, 切	13	2	20	7	194	3.6%
12	キョウチョウ・スル	協調, 強調	16	2	20	150	195	76.9%
13	ケントウ・スル	検討, 健闘	25	2	28	314	344	91.3%
14	コウセイ・スル	構成, 校正	15	2	18	125	328	38.1%
15	コエ・ル	越え, 超え	18	2	25	19	498	3.8%
16	コタエ・ル	答え, 応え	16	2	19	81	159	50.9%
17	サ・ス	指, 刺,	12	2	16	27	182	14.8%
18	サケ・ル	避け, 裂け	14	2	16	228	236	96.6%
19	サンカ・スル	酸化, 参加	20	2	22	349	426	81.9%
20	シボウ・スル	死亡, 志望	8	2	8	54	120	45%
21	シメ・ル	占め, 絞め, 締 め, 閉め	16	4	26	31	242	12.8%
22	ススメ・ル	進め, 勧め	32	2	45	38	310	12.3%
23	セイサン・スル	生産, 清算	14	2	17	64	111	57.7%
24	センタク・スル	選択, 洗濯	22	2	25	63	145	43.4%
25	ソナエ・ル	備え, 供え	15	2	18	18	235	7.7%
26	タ・ツ	た, 立, 断	23	3	29	317	425	74.6%
27	タズネ・ル	尋ね, 訪ね	16	2	26	4	176	2.3%
28	ツ・ク	着, 付, 就, 突	21	4	27	30	243	12.3%
29	ツレ・ル	連れ, 釣れ	19	2	22	27	108	25%
30	トウジョウ・スル	登場, 搭乗	12	2	16	26	148	17.6%
31	ニ・ル	似, 煮	19	2	20	225	241	93.4%
32	ノ・ル	載, 乗	15	2	24	15	276	5.4%
33	ノゾ・ク	覗, 除	6	2	8	14	234	6%
34	ハカ・ル	図, 計量, 諮	15	4	21	49	313	15.7%
35	ハッセイ・スル	発声, 発生	15	2	18	136	343	39.7%
36	ハナ・ス	話, 離	30	2	37	81	287	28.2%
37	ヒロゲ・ル	拡げ, 広げ	18	2	20	165	193	85.5%
38	ヘ・ル	減, 経	21	2	21	324	327	99.1%
39	ヘンカン・スル	変換, 返還	20	2	26	24	144	16.7%
40	ミエ・ル	みえ, 見え	23	2	26	54	427	12.6%
41	ヨウセイ・スル	要請, 養成	16	2	20	61	427	14.3%

表 15 京大コーパスから抽出した「あう」の格助詞部分

グループ ID	表記	格助詞	格助詞	格助詞	格助詞	出典	例文
1	会う	と(カストロ(名詞) 国家(名詞) 評議会(名詞) 議長(名詞))				950101044- 001	サッカー界のスーパースター、ディエゴ・マラドーナ選手とその家族が昨年 12 月 30 日、キューバのカストロ国家評議会議長と会い、ニコリ記念撮影。
1	会う	で(東京(名詞) 駅(名詞))				950101118- 003	二十三日、東京都千代田区の交通博物館で長坂容疑者に声を掛けられ、二十六日、長坂容疑者からの電話で東京駅で会う約束をした。
1	会う	に(山岳(名詞) 民族(名詞))				950101160- 003	デルタの農民、カンボジアのチャム族、ラオス、ミャンマーの山岳民族にも会い、文化遺産を継承する人々とメコン川の将来について語り合った。
1	会う	と(人(名詞) たち(接尾辞))				950101161- 023	「開発の波がラオスやミャンマー、中国に拡大している。開発のゆがみを知る人々は、そうした国の人たちと会い、経験を伝える必要がある」。

合は、該当パターンなしで×とする。同じ読みの単位で、

正解の数/事例数

を正解率とする。評価は①の段階で判定できた割合、②の段階まで入れて判定できた割合を出している。この正解率を 41 組の各組ごとに出した結果を表 17 に示す。

なお、格助詞パターンで複数解の出るものには、パターンの頻度を導入することによって正解になる場合と、パターンの頻度を導入しても正解に至らない場合とがある。

全体的な基準パターンとの照合結果は以下ようになった。

- ① 基準パターン表において格助詞パターンに相当する表記が単一でそれを表記にわりあてた結果、正解になったもの：1604 件 (39.9%)
- ② 基準パターン表にパターンそのものがなかった、あるいはパターンはあったが判定結果が誤りにあったもの：2420 件 (61.1%)
- ③ ②のうち、パターンは合致したが、第 1 段階で判定結果誤りとなったものについて同音異表記セット内のパターンの出現頻度を導入し、その大小で決定できたもの：1,316 件 (32.70%)

以上の結果、第 1 段階として格助詞パターンの違い、第 2 段階として格助詞パターンの頻度の大小を導入して、格助詞パターンのレベルだけを用いたテストデータの判定で 72.7%が正解となった。

表 16 「あう」に関して京大コーパスを対象とした判定

読み が同 じも のの 番号	京大コー パスの正 解表記	① パターン判 定表で複数 候補	× 表記 が複数 候補あ るもの	② コーパス の同表記 内の出現 頻度の大小で判定 した表記	頻度で 判定し た表記 が判定 誤りだ ったもの	格助詞組 み合わせ	格助詞	格助詞
1	会う	会う, 合う	×	会う		_DE	で(現地(名詞))	
1	会う	会う, 合う	×	会う		_DE	で(取材(名詞))	
1	会う	会う, 合う	×	会う		_DE	で(場(名詞))	
1	会う	会う, 合う	×	会う		_DE	で(東京(名詞) 駅(名詞))	
1	会う	会う, 合う	×	会う		_DE	で(料亭(名詞))	
1	合う	会う, 合う	×	会う	×	_DE	で(政策(名詞) 立案(名詞))	
1	会う	会う		会う		_DE_TO	で(こと(名詞))	と(二(名詞)人 (接尾辞))
1	会う	会う		会う		_DE_TO	で(デトロイト (名詞))	と(首脳(名詞))
1	会う	会う		会う		_DE_TO	で(場(名詞))	と(江(名詞)沢 民(名詞)国家 (名詞)主席(名 詞))
1	会う	会う		会う		_DE_TO	で(都内(名詞))	と(海江田(名 詞)万里(名詞) 衆院(名詞)議 員(名詞))
1	会う	会う		会う		_DE_TO	と(主婦(名詞))	で(ほとり(名 詞))
1	会う	会う		会う		_DE_TO	と(方々(名詞))	で(水面(名詞) 下(名詞))
1	合う	会う, 合う	×	合う		_GA	が(「(特殊)目 (名詞))	
1	合う	会う, 合う	×	合う		_GA	が(コンビ(名 詞))	
1	合う	会う, 合う	×	合う		_GA	が(各国(名詞))	
1	合う	会う, 合う	×	合う		_GA	が(靴(名詞))	
1	合う	会う, 合う	×	合う		_GA	が(兄弟(名詞) 民族(名詞))	
1	合う	会う, 合う	×	合う		_GA	が(経営者(名 詞))	
1	合う	会う, 合う	×	合う		_GA	が(国民(名詞))	
1	合う	会う, 合う	×	合う		_GA	が(心(名詞))	
1	合う	会う, 合う	×	合う		_GA	が(文部省(名 詞))	
1	合う	会う, 合う	×	合う		_GA	が(目(名詞))	
1	合う	会う, 合う	×	合う		_GA	が(隣人(名詞) 同士(接尾辞))	

表 17 京大コーパスを対象にした, 格助詞組み合わせによる異表記判定の結果

	読み	表記	グループ 内表記数	パターンが一致して判定できた場合の正解率	パターン頻度の大小を参照して判定できた場合の正解率
1	ア・ウ	会, 合, 遭	3	12%	35.3%
2	アガ・ル	上が, 拳が	2	42.5%	72.5%
3	アラワ・ス	表わ, 著	2	33.3%	33.3%
4	イ・ク	逝, 行	2	64%	72%
5	ウツ・ル	映, 写, 移	3	22.8%	49.4%
6	オク・ル	送, 贈	2	8.3%	69.7%
7	カ・ウ	買, 飼	2	10.3%	80.9%
8	カ・ク	書, 欠, 描	3	23.5%	49.7%
9	カエ・ル	変え, 帰, 買え, 替え, 代え	5	10.9%	76.3%
10	カンセイ・スル	完成, 管制	2	65.2%	87%
11	キ・ル	着, 切	2	10.5%	36.8%
12	キョウチョウ・スル	協調, 強調	2	48.3%	56.2%
13	ケントウ・スル	検討, 健闘	2	74%	82.1%
14	コウセイ・スル	構成, 校正	2	77.8%	88.9%
15	コエ・ル	越え, 超え	2	6.1%	79.6%
16	コタエ・ル	答え, 応え	2	42.1%	54.3%
17	サ・ス	指, 刺	2	20.6%	44.1%
18	サケ・ル	避け, 裂け	2	86.1%	86.1%
19	サンカ・スル	酸化, 参加	2	64.1%	87.6%
20	シボウ・スル	死亡, 志望	2	76.7%	76.7%
21	シメ・ル	占め, 絞め, 締め, 閉め	4	20.8%	71.5%
22	ススメ・ル	進め, 勧め	2	17.7%	75.8%
23	セイサン・スル	生産, 清算	2	38.1%	52.4%
24	センタク・スル	選択, 洗濯	2	16.7%	91.7%
25	ソナエ・ル	備え, 供え	2	87.3%	87.3%
26	タ・ツ	た, 立, 断	3	63.5%	75%
27	タズネ・ル	尋ね, 訪ね	2	4.3%	52.2%
28	ツ・ク	着, 付, 就, 突	4	18%	37.7%
29	ツレ・ル	連れ, 釣れ	2	14.3%	66.7%
30	トウジョウ・スル	登場, 搭乗	2	63.6%	76.4%
31	ニ・ル	似, 煮	2	94.1%	94.1%
32	ノ・ル	載, 乗	2	7.7%	75.8%
33	ノゾ・ク	覗, 除,	2	3.6%	94%
34	ハカ・ル	図, 計, 量, 諮	4	17%	86.2%
35	ハッセイ・スル	発声, 発生	2	74.7%	94.7%
36	ハナ・ス	話, 離	2	83.5%	90%
37	ヒロゲ・ル	広げ, 広げ	2	39.5%	85.2%
38	ヘ・ル	減, 経	2	90.1%	90.1%
39	ヘンカン・スル	変換, 返還	2	41.7%	8.3%
40	ミエ・ル	みえ, 見え	2	41.5%	72.6%
41	ヨウセイ・スル	要請, 養成	2	36.2%	68.1%

5 最後に

以上、本研究では動詞概念約 12,400 について、約 15,700 の事例から、動詞の結合価の第一レベルに相当する格助詞組み合わせ範囲に着目した検討を行った。上記範囲で格助詞組み合わせの延べパターンは約 37,000 パターン、異なりパターンにして 188 パターンに集約された。またこれらの格助詞のみのパターンから、重要と思われる格助詞からおおまかに動詞グループをまとめると、①対象格をとる動詞群が約 55 %、②「が格」をとる現象動詞の群が約 18.5 %、③GOAL の「に格」を採る動詞が 16.9 %で、これだけで動詞全体の約 91 %を占めることがわかった。

また、同音異表記に該当する 41 セットについて、評価データの格助詞パターンがどれぐらいまで判別可能かを検討したところ、格助詞パターンのみの判定で 39.9 %、異表記セット内での格助詞パターンの出現率も考慮して、平均 72.7 %の判定であった。

以上、これらの数量的検討は延べにして約 157,000 の動詞概念を含む、品詞などの形態素情報や係り受けなどの構文情報を備えた EDR コーパスを元にして可能となった大量の言語事象の分析および研究である。

また、「日本語動詞の結合価」データには、EDR コーパスに出現した体言が格助詞別に配置され実例の形で掲載されている。これは、体言部分の意味分析にも貴重な言語データである。現在この体言部分の意味分類についても研究中であるが、これは機会を改めて述べたい。

謝辞

本研究の元となった「日本語動詞の結合価」は、以下の諸氏による綿密なチェックを経て、充実した言語データとなったものである。大久保佳子(日本システムアプリケーション)、丸元聡子(計量計画研究所)、衛藤純司(ランゲージウェア)、大島玲子氏らに感謝する。

また、本研究の評価を行うために京大コーパスから該当部分のデータ抽出および形式化の部分について、内元清貴氏(独立行政法人 情報通信研究機構)にご協力いただいた。合わせて謝意を表する。

参考文献

- 相場徹, 奥村学 (1995). “構文・意味解析と統合した形態素解析に関する研究.” EDR 電子化辞書利用シンポジウム論文集, pp. 41-88.
- 林良彦 (1987). “結合価構造に基づく日本文解析.” 情報処理学会研究報告 87-NL-62, pp. 39-42.
- 平岡冠二, 松本裕治 (1994). “コーパスからの動詞の格フレーム獲得と名詞のクラスタリング.” 情報処理学会研究報告 104-NL-11, pp. 79-86.
- 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩己, 小倉健太郎, 大山芳史, 林良彦 (1997). 日

本語彙大系. 岩波書店.

石綿敏雄 (1975). “日本語の生成語彙論的記述と言語処理への応用.” 電子計算機による国語研究 VII. 国立国語研究所報告 54. 秀英出版.

情報処理振興事業協会 (1987). 計算機用日本語基本動詞辞書 IPAL 解説編. 情報処理振興事業協会.

鍛冶伸裕, 河原大輔, 黒橋禎夫, 佐藤理史 (2003). “格フレームの対応付けに基づく用言の言い換え.” 自然言語処理, 10 (4), pp. 65–81.

金出地真人, 徳久雅人, 村上仁一, 池原悟 (2003). “結合価文法による動詞と名詞の訳語選択能力の評価.” 情報処理学会研究報告 2003-NL-153, pp. 119–124.

河原大輔, 黒橋禎夫 (2002). “用言と直前の格要素の組を単位とする格フレームの自動獲得.” 自然言語処理, 9 (1), pp. 3–19.

国立国語研究所 (1997). “日本語における表層格と深層格の対応関係.” 国立国語研究所報告 113. 三省堂.

峯恒憲, 實成暁司, 中山聡, 久保山哲二, 谷口倫一郎, 雨宮真人 (1995). “動詞語義の分類法.” EDR 電子化辞書利用シンポジウム論文集, pp. 89–96.

水谷静夫, 石綿敏雄, 荻野孝野, 賀来直子, 草薙裕 (1983). 朝倉日本語新講座 3 文法と意味 1. 朝倉書店.

日本電子化辞書研究所 (1995). EDR 電子化辞書仕様説明書.

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, 吉田紀子, 小林正博, 井佐原均 (2005). “日本語動詞の結合価記述に必要な意味分類項目と意味分類項目数.” 計量国語学, 25 (1). (掲載予定).

白井諭, 横尾昭男, 中岩浩巳 (1997). “日英機械翻訳のための構文辞書.” 情報処理学会研究報告 91-NL-84, pp. 95–102.

付録

A 「会う」の結合価データ

会う アウ 3ce8ca 会う										その他	例文
は	が	を	に	へ	から	より	まで	で	と		
	が (市民 (名詞))		に (記者 (名詞))							'@renyou (な (形容 詞))	一般市民が、こた わりもなく西側の 記者に [[会]] うこ とは、やはりまだ できない。
								で (室 (名 詞))		か (何度 (名詞))	編集長室で何度が [[会]] ったが、いつ もタートルネック のセーターにぎっ くりした背広をは おっていた。
			は (少女 (名詞))					で (どこ (名詞))		に (前 (名 詞))	今、すれ違った短 い髪の少女は、前 にどこかで [[会]] ったような気がす る。
	が (委員 長 (名詞))		に (政治 家 (名詞))					で (某所 (名詞))			新政連中央委員会 の小林永司副委員 長が、某所で 17、 8 人の政治家に [[会]] った。
	が (会長 (名詞))		に (通産 大臣 (名 詞))								昨年末の石炭の問 題でもね、武田豊 日本鉄鋼連盟会長 が最終的には通産 大臣に [[会]] って、 全面協力すること になりました。
	が (彼 (名 詞))							で (2 人 (複合語))	と (首相 (名詞))		彼が首相と 2 人 きりで [[会]] った のは、今年の記録 を見る限り、この 時の 1 回だけであ る。
	が (国民 (名詞))										米国滞在中に米国 民がまだ [[会]] っ たことのない人に 会い、好感を抱く ことを知と思う。

B 格助詞パターン別単語事例

格助詞パターン	概念異なり数	事例頻度合計	動詞(例)
_WO	4502	32414	ア・ム [編む](1) ア・ム [編む](2) アイ・スル [愛する](1) アイ・スル [愛する](1) アイ・スル [愛する](1) アイ・スル [愛する](5) アイ・スル [愛する](2) アイテド・ル [相手取る](1) アイドク・スル [愛読する](1) アイヨウ・スル [愛用する](1) (1) :
_GA_WO	3820	23868	ア・ム [編む](8) アイ・スル [愛する](3) アイ・スル [愛する](2) アイ・スル [愛する](3) アイ・スル [愛する](7) アイ・スル [愛する](1) アイコウ・スル [愛好する](1) アイテド・ル [相手取る](3) アイヨウ・スル [愛用する](7) アオ・グ [仰ぐ](1) (1) :
_GA	4016	22141	ア・ウ [会う](11) ア・ウ [合う](16) ア・ウ [合う](1) ア・ウ [合う](2) ア・ウ [合う](5) ア・ウ [合う](2) ア・ク [聞く](5) ア・ク [聞く](1) ア・ク [空く](2) ア・ク [空く](1) (1) :
_GA_NI	2270	13352	ア・ウ [会う](31) ア・ウ [合う](11) ア・ウ [合う](8) ア・ウ [合う](8) ア・ウ [遭う](16) ア・ル [ある](94) ア・ル [ある](6) ア・ル [ある](500) アイツ・グ [相次ぐ](12) アイノリ・スル [相乗りする](3) (3) :
_NI	1804	9031	ア・ウ [会う](48) ア・ウ [合う](25) ア・ウ [合う](3) ア・ウ [合う](1) ア・ウ [合う](8) ア・ウ [合う](4) ア・ウ [遭う](8) ア・ル [ある](6) ア・ル [ある](1) ア・ル [ある](28) (28) :
_WO_NI	1694	8624	ア・ム [編む](1) アイノリ・スル [相乗りする](1) アオ・グ [仰ぐ](5) アカ・ス [明かす](1) アケ・ル [開ける](7) アケワ・ス [明け渡す](1) アゲ・ル [挙げる](3) アゲ・ル [挙げる](7) アゲ・ル [上げる](1) アゲ・ル [上げる](5) (5) :
_WO_DE	2058	7965	ア・ム [編む](3) アイヨウ・スル [愛用する](1) アオ・グ [仰ぐ](1) アカ・ス [明かす](3) アカラメ・ル [赤らめる](1) アクヨウ・スル [悪用する](1) アケ・ル [開ける](1) アゲ・ル [挙げる](2) アゲ・ル [上げる](1) アゲ・ル [上げる](6) (6) :
_GA_DE	1739	5566	ア・ウ [会う](7) ア・ウ [合う](1) ア・ク [空く](1) ア・ル [ある](11) ア・ル [ある](24) アイツ・グ [相次ぐ](43) アカ・ス [明かす](1) アガ・ル [上がる](11) アガリコ・ム [上がり込む](1) アクシュ・スル [握手する](1) (1) :
格助詞ナシ	2086	4915	ア・ウ [会う](11) ア・ウ [合う](2) ア・ウ [合う](4) ア・ウ [合う](1) ア・ク [空く](1) ア・ル [ある](2) アイ・スル [愛する](1) アイツ・グ [相次ぐ](1) アイヨウ・スル [愛用する](1) アオ・ル [あおる](1) (1) :
_GA_WO_NI	1003	3222	アイヨウ・スル [愛用する](1) アオ・グ [仰ぐ](1) アカ・ス [明かす](1) アケ・ル [開ける](1) アケワ・ス [明け渡す](1) アゲ・ル [挙げる](2) アゲ・ル [上げる](1) アゲ・ル [上げる](2) アゲ・ル [上げる](1) アズカ・ル [預かる](1) (1) :
_GA_WO_DE	1223	3173	アカ・ス [明かす](3) アゲ・ル [挙げる](1) アゲ・ル [上げる](1) アゲ・ル [上げる](3) アゲ・ル [上げる](1) アタメ・ル [暖める](1) アットウ・スル [圧倒する](2) アツパク・スル [圧倒する](1) アツカ・ウ [扱う](1) アツカ・ウ [扱う](1) (1) :
_DE	1310	3001	ア・ウ [会う](12) ア・ウ [合う](1) ア・ル [ある](2) アイ・スル [愛する](1) アイツ・グ [相次ぐ](1) ア・ル [あおる](2) アガ・ル [上がる](1) アガ・ル [上がる](1) アガナ・ウ [あがなう](1) アキラメ・ル [諦める](1) (1) :
_TO	366	2079	ア・ウ [会う](31) アイハン・スル [相反する](1) アクシュ・スル [握手する](4) アソ・ブ [遊ぶ](1) アツマ・ル [集まる](1) アワセ・ル [合わせる](1) アイウ・スル [併せる](13) アワセ・ル [併る](1) アンナイ・スル [案内する](1) イ・ウ [いう](240) (240) :
_WO_KR	611	1501	アイドク・スル [愛読する](1) アゲ・ル [挙げる](1) アゲ・ル [上げる](3) アズカ・ル [預かる](2) アタエ・ル [与える](1) アツカ・ウ [扱う](1) アツメ・ル [集める](17) アツメ・ル [集める](2) アテ・ル [当てる](1) アテ・ル [当てる](1) (1) :
_GA_TO	413	1485	ア・ウ [会う](25) ア・ウ [合う](1) ア・ウ [合う](10) ア・ウ [合う](1) ア・ウ [合う](2) ア・ル [ある](1) ア・ル [ある](1) アイセンゴ・スル [相前後する](1) アイハン・スル [相反する](1) アイマツ・ツ [相まつ](1) (1) :
_GA_KR	477	1287	ア・ル [ある](5) ア・ル [ある](1) アイツ・グ [相次ぐ](8) アガ・ル [上がる](1) アガ・ル [上がる](2) アケ・ル [開ける](3) アツカ・スル [悪化する](2) アツカエ・ル [扱える](1) アツマ・ル [集まる](5) アツマ・ル [集まる](11) (11) :
_WO_TO	219	1247	アオ・グ [仰ぐ](3) アガメ・ル [あがめる](1) アテ・ル [当てる](1) アナド・ル [悔める](1) アユ・ム [歩む](1) アラタメ・ル [改める](1) アル・ク [歩く](3) アルキマワ・ル [歩き回る](1) アワセ・ル [合わせる](3) アワセ・ル [合わせる](4) (4) :
_GA_NI_DE	492	939	ア・ウ [会う](10) ア・ウ [合う](1) ア・ウ [遭う](5) ア・ル [ある](1) ア・ル [ある](2) ア・ル [ある](8) アカ・ス [明かす](1) アガ・ル [上がる](2) アクセス・スル [アクセスする](1) アセ・ル [焦る](1) (1) :

C 格助詞異なりパターン表

格数	格名 1	格名 2	格名 3	格名 4	格名 5	格名 6	頻度	第 1 分類	第 2 分類
1	WO						32414	WO	
2	GA	WO					23868	WO	
2	WO	DE					7965	WO	
3	GA	WO	DE				3173	WO	
2	HA	WO					853	WO	
3	HA	GA	WO				171	WO	
3	HA	WO	DE				152	WO	
4	HA	GA	WO	DE			24	WO	
2	WO	NI					8624	WO	NI
3	GA	WO	NI				3222	WO	NI
3	WO	NI	DE				658	WO	NI
3	WO	NI	KR				222	WO	NI
4	GA	WO	NI	DE			188	WO	NI
2	WO	HE					163	WO	NI
3	HA	WO	NI				142	WO	NI
3	GA	WO	HE				79	WO	NI
4	GA	WO	NI	KR			42	WO	NI
3	WO	HE	KR				39	WO	NI
3	WO	HE	DE				18	WO	NI
4	WO	NI	KR	DE			17	WO	NI
4	HA	GA	WO	NI			12	WO	NI
3	WO	NI	TO				12	WO	NI
4	GA	WO	HE	KR			10	WO	NI
4	GA	WO	NI	TO			7	WO	NI
4	WO	NI	KR	MD			7	WO	NI
4	WO	NI	DE	TO			6	WO	NI
4	GA	WO	HE	DE			5	WO	NI
4	HA	WO	NI	DE			5	WO	NI
3	WO	NI	HE				5	WO	NI
3	HA	WO	HE				4	WO	NI
4	WO	HE	KR	DE			4	WO	NI
4	HA	WO	NI	KR			3	WO	NI
3	WO	NI	YR				3	WO	NI
5	GA	WO	NI	KR	DE		2	WO	NI
4	GA	WO	NI	YR			2	WO	NI
4	GA	WO	NI	MD			2	WO	NI
4	HA	WO	HE	KR			2	WO	NI
4	GA	WO	HE	MD			1	WO	NI
4	GA	WO	NI	HE			1	WO	NI
5	GA	WO	NI	MD	DE		1	WO	NI
5	GA	WO	NI	DE	TO		1	WO	NI
5	HA	GA	WO	NI	DE		1	WO	NI
4	HA	WO	NI	TO			1	WO	NI
4	WO	NI	HE	MD			1	WO	NI
4	WO	NI	KR	TO			1	WO	NI
4	WO	NI	YR	DE			1	WO	NI
2	WO	KR					1501	WO	KR
3	GA	WO	KR				606	WO	KR
3	WO	KR	DE				147	WO	KR
4	GA	WO	KR	DE			51	WO	KR
2	WO	YR					39	WO	KR
3	HA	WO	KR				32	WO	KR
3	GA	WO	YR				22	WO	KR
3	WO	KR	MD				22	WO	KR
4	GA	WO	KR	MD			11	WO	KR

D 同じ概念内の格助詞パターングループと基本格

基本格サイン * グループ内の基本格パターン

概念グループ番号	概念グループ内パターングループ番号	基本格サイン	格助詞パターン	頻度	同概念内割合	概念	概念説明
12	1			27	アガ・ル [上がる]	0ee91b	終わりになる
12	1		_GA_KR	1	アガ・ル [上がる]	0ee91b	終わりになる
12	1		_KR	2	アガ・ル [上がる]	0ee91b	終わりになる
12	1		_GA	6	アガ・ル [上がる]	0ee91b	終わりになる
13	1		_GA_NI	10	アガ・ル [上がる]	3cf546	前からあるものに別のものを合わせて程度を増す
13	1		_GA_NI_DE	2	アガ・ル [上がる]	3cf546	前からあるものに別のものを合わせて程度を増す
13	1	*	_GA_NI_KR	1	アガ・ル [上がる]	3cf546	前からあるものに別のものを合わせて程度を増す
13	1		_HA_GA_NI	2	アガ・ル [上がる]	3cf546	前からあるものに別のものを合わせて程度を増す
13	1		_NI_KR_DE	1	アガ・ル [上がる]	3cf546	前からあるものに別のものを合わせて程度を増す
13	2	*	_GA_MD	5	アガ・ル [上がる]	3cf546	前からあるものに別のものを合わせて程度を増す
13	2		_MD	2	アガ・ル [上がる]	3cf546	前からあるものに別のものを合わせて程度を増す
13	2		_MD_DE	1	アガ・ル [上がる]	3cf546	前からあるものに別のものを合わせて程度を増す
13	2		_GA_YR	2	アガ・ル [上がる]	3cf546	前からあるものに別のものを合わせて程度を増す
13	12		_DE	1	アガ・ル [上がる]	3cf546	前からあるものに別のものを合わせて程度を増す
13	12		_GA	105	アガ・ル [上がる]	3cf546	前からあるものに別のものを合わせて程度を増す
13	12		_GA_DE	11	アガ・ル [上がる]	3cf546	前からあるものに別のものを合わせて程度を増す
13	12		_GA_KR	2	アガ・ル [上がる]	3cf546	前からあるものに別のものを合わせて程度を増す
13	12		_HA_GA	5	アガ・ル [上がる]	3cf546	前からあるものに別のものを合わせて程度を増す
13	12			453	アガ・ル [上がる]	3cf546	前からあるものに別のものを合わせて程度を増す

略歴

荻野 孝野 (正会員): 1971 年東京女子大学文学部卒業。同年 (財) 計量計画研究所言語情報研究室入所。1986 年 日本電子化辞書研究所にて EDR 電子化辞書開発に従事。現在, (株) 日本システムアプリケーション専任顧問, 明海大学非常勤講師。2000 年神戸大学自然科学研究科博士課程在籍。自然言語処理, 主として概念体系, 結合価の研究にかかわる。言語処理学会, 人工知能学会, 情報処理学会, 計量国語学会, 各会員。

植田 禎子: 2001 年法政大学文学部卒業. 2003 年東京工業大学院総合理工学研究科修士課程修了. 現在, (株) 日本システムアプリケーション.

小林 正博: 1978 年東京理科大学卒業. 同年日本システムアプリケーション入社. 1992 年から (株) 日本電子化辞書研究所出向にて EDR 電子化辞書開発の計算機部門を担当. 現在, (株) 日本システムアプリケーション. 情報処理学会会員.

井佐原 均 (正会員): 1980 年京都大学大学院修士課程修了. 博士 (工学). 同年通商産業省電子技術総合研究所入所. 1995 年郵政省通信総合研究所. 現在, 独立行政法人情報通信研究機構けいはんな情報通信融合研究センター自然言語グループリーダーおよびタイ自然言語ラボラトリー長. および, 神戸大学大学院自然科学研究科教授. 自然言語処理, 語彙意味論の研究に従事. 言語処理学会, 情報処理学会, 人工知能学会, 日本認知科学会, ACL, 各会員.

(2004 年 9 月 14 日 受付)

(2004 年 12 月 08 日 再受付)

(2005 年 4 月 23 日 採録)