

## 第7章

# 日本語共起辞書

日本語共起辞書は、日本語コーパスに格納された実例文の解析結果から、係り受けを構成している部分、すなわち共起句を抽出し、句の表記の五十音順に並べたものである。日本語共起辞書レコードは、レコード番号、見出し情報、共起句構成要素情報、構文情報、意味情報、共起状況情報、および管理情報から構成される。

日本語共起辞書の基本的な役割は、日本語コーパス中での共起状況の情報に基づいて妥当な自立語の組合せ方の実例を示すことである。

日本語共起辞書には、主要な動詞について表層および深層の格に関連する情報を記述した、日本語動詞共起パターン副辞書が付属する。

===== [日本語共起辞書レコードの構造] =====

<レコード番号> *1	: レコードタイプと識別番号
<見出し情報>	: 共起辞書レコードの見出し(→7.1節)
<句見出し> *2	: 共起句の表記
<共起句構成要素情報>	: 共起句を構成する形態素列に関する情報(→7.2節)
<構成要素列> *3	: 共起句を構成する個々の形態素
<構文情報>	: 共起句の構文構造を示す情報(→7.3節)
<部分構文木> *4	: 共起句の構造を示す構文木
<意味情報>	: 深層の概念関係を示す情報(→7.4節)
<部分意味フレーム> *5	: 概念関係を示す意味フレーム情報
<共起状況情報>	: 日本語コーパスにおける共起状況を示す情報(→7.5節)
<頻度> *6	: 日本語コーパスにおける出現頻度(→7.5.1節)
<例文>... *7	: 日本語コーパスで出現した文を共起関係に基づき再構成した文(→7.5.2節)
<管理情報>	: 辞書開発の管理のための情報
<管理履歴レコード>*8	: 更新日付等の管理情報

\*1

<レコード番号> ::= JCC <識別番号>

-----

<識別番号> : 7桁の10進数

\*2

<句見出し> ::= <単語表記1>¥t<共起関係子>¥t<単語表記2>

<単語表記1> ::= <単語表記>

<単語表記2> ::= <単語表記>

-----

<単語表記> : 共起句を構成する単語の表記文字列

<共起関係子> : 表7-2に示す共起関係子

## \*3

〈構成要素列〉	::= 〈構成要素〉。。
〈構成要素〉	::= ' { 〈要素番号〉 〈形態素〉 〈かな表記〉 〈品詞〉 〈慣用句フラグ〉 〈概念情報〉 } '
〈概念情報〉	::= ""   〈意味情報〉   〈補足付き概念説明〉
〈意味情報〉	::= 〈概念識別子〉 〈概念見出し〉 〈概念説明〉
〈概念識別子〉	::= 〈16進整数〉
〈概念見出し〉	::= 〈英語概念見出し〉 〈日本語概念見出し〉
〈概念説明〉	::= 〈英語概念説明〉 〈日本語概念説明〉
〈補足付き概念説明〉	::= " 〈文字列〉 "

〈要素番号〉	: 文中の出現順に従った通し番号
〈形態素〉	: 共起句を構成する形態素の表記
〈かな表記〉	: 共起句を構成する形態素のかな表記
〈品詞〉	: 共起句を構成する形態素の品詞
〈慣用句フラグ〉	: 形態素の概念との対応関係の種類。 : (0)通常概念であり直接対応する。 : (1)慣用句の一部であり直接対応しない。
〈日本語概念見出し〉	: 概念の日本語概念見出し
〈英語概念見出し〉	: 概念の英語概念見出し
〈日本語概念説明〉	: 概念の日本語による説明
〈英語概念説明〉	: 概念の英語による説明
〈複合概念説明〉	: 複合概念の日本語による説明
〈補足付き概念説明〉	: 単語辞書の単の概念内に適当な概念がない場合に記述される概念の説明

## \*4

〈部分構文木〉	::= 〈受け側要素〉 〈関係要素〉 〈係り側要素〉
〈受け側要素〉	::= 〈受け側要素番号〉 / 〈受け側単語表記〉
〈関係要素〉	::= 〈関係要素番号〉 / 〈共起関係子〉 / 〈関係単語表記〉
〈係り側要素〉	::= 〈係り側要素番号〉 / 〈係り側単語表記〉

〈受け側要素番号〉	: 受け側に対応する要素番号のリスト
〈受け側単語表記〉	: 受け側の単語表記
〈関係要素番号〉	: 共起関係に対応する要素番号のリスト
〈共起関係子〉	: 表7-2に示す共起関係子
〈関係単語表記〉	: 共起関係に対応する単語表記
〈係り側要素番号〉	: 係り側に対応する要素番号のリスト
〈係り側単語表記〉	: 係り側の単語表記

## \*5

〈部分意味フレーム〉	::= 〈受け側概念要素〉¥t{ ""   〈概念関係子〉 }¥t〈係り側概念要素〉
〈受け側概念要素〉	::= ""   〈受け側要素番号〉 / 〈受け側概念識別子〉 / 〈受け側単語表記〉
〈係り側概念要素〉	::= ""   〈係り側要素番号〉 / 〈係り側概念識別子〉 / 〈係り側単語表記〉

〈受け側要素番号〉	: 受け側に対応する要素番号のリスト
〈受け側概念識別子〉	: 受け側の概念識別子
〈受け側単語表記〉	: 受け側の単語表記

<概念関係子> : 深層の概念関係を示す関係子  
 <係り側要素番号> : 係り側に対応する要素番号のリスト  
 <係り側概念識別子> : 係り側の概念識別子  
 <係り側単語表記> : 係り側の単語表記

\*6

<頻度> ::= <表層共起頻度> ; <共起項目頻度> ; <受け側形態素頻度> ;  
 <係り側形態素頻度>

-----  
 <表層共起頻度> : 表層の共起関係の出現頻度  
 <共起項目頻度> : 深層の概念関係を含む出現頻度  
 <受け側形態素頻度> : 受け側形態素の頻度  
 <係り側形態素頻度> : 係り側形態素の頻度

\*7

<例文> ::= ' { <文識別子情報> / <文字列> } '  
 <文識別子情報> ::= <文識別子> ( ; ) 。 。 。  
 <文識別子> ::= <文字列>

\*8

<管理履歴レコード> ::= <属性名> = <属性値> | <管理履歴レコード> ; <属性名> = <属性値>  
 <属性名> ::= <文字列>  
 <属性値> ::= <文字列>

=====

```

===== [日本語共起辞書レコードの例] =====
<レコード番号>                                JCC7173641
<見出し情報>
    <句見出し>                                昼食 を 食べ
<共起句構成要素情報>
<要素  <形態  <かな表記  <品詞> <慣用句 <概念情報>
 番号> 素>                                フラグ>
{ 1    昼食   チュウシヨク  名詞   0      3bec74  lunch  昼食 [チュウシヨク]  "a
meal eaten at noon"  昼の食事 }
{ 2    を     ヲ             助詞   0      "" }
{ 3    食べ   タベ         動詞   0      3bc6f0  ""  食べる [タベ・ル]  "to eat
something"  食物をとる }
<構文情報>
    <部分構文木>
        <受け側要素>                        3/食べ
        <関係要素>                          2/を/を
        <係り側要素>                        1/昼食
<意味情報>
    <部分意味フレーム>
        <受け側概念要素>                    3/3bc6f0/食べ
        <概念関係子>                        object
        <係り側概念要素>                    1/3bec74/昼食
<共起状況情報>
    <頻度>                                    1;1;488;6
    <例文>                                    {00050003b57d-8-3/<昼食>を…(食べ)に帰る}
<管理情報>
    <管理履歴レコード>                      DATE="95/3/31"

```

```

===== [日本語共起辞書レコードの例] =====
<レコード番号>                                JCC5321382
<見出し情報>
    <句見出し>                                借り @rentai 本
<共起句構成要素情報>
<要素  <形態  <かな  <品詞> <慣用句 <概念情報>
 番号> 素>    表記>                                フラグ>
{ 1    借り   カリ       動詞   0      3cfdb4  borrow  借りる [カリ・ル]  "to use a
person's property after promising to give it back to the lender"  返す約束で、他人のものを
使う }
{ 2    た     タ         助動詞 0      "" }
{ 3    本     ホン      名詞   0      0e5097  volume  本 [ホン]  "Publications" 書籍 }
<構文情報>
    <部分構文木>
        <受け側要素>                        1/借り
        <関係要素>                          2/@rentai/た
        <係り側要素>                        3/本
<意味情報>

```

〈部分意味フレーム〉	
〈受け側概念要素〉	1/3cfdb4/借り
〈概念関係子〉	
〈係り側概念要素〉	3/0e5097/本
〈共起状況情報〉	
〈頻度〉	5;1;562;231
〈例文〉	{000600000067-18-15/<借り>た(本)}
〈管理情報〉	
〈管理履歴レコード〉	DATE="95/3/31"

=====

## 7.1 見出し情報

日本語共起辞書の見出し情報は、共起句を構成する受け側および係り側の形態素の表記と共起関係子（→7.3節）を連結した句見出しである。連結する順番は、受け側・係り側の形態素で文中に始めに登場するもの、共起関係子、受け側・係り側の形態素で文中に後ろに登場するものの順である。

日本語共起辞書に記述する係り受けの種類は、格関係、用言・連体詞と体言との連体修飾関係、格助詞「の」を介した体言間の連体修飾関係、用言・副詞と体言との連用修飾関係、助数詞による体言の修飾関係の5種類である。

## 7.2 共起句構成要素情報

共起句構成要素情報は、共起句に対応する実例文における形態素の表記・かな表記・品詞・慣用句フラグ・概念情報を対にして並べたものである。ここで、受け側・係り側・共起関係子に対応する形態素がすべて示される。

かな表記には、形態素の読みをカタカナで記述する。形態素が数字である場合は、算用数字で表記する。

品詞は、日本語単語辞書の品詞を名詞・動詞などに大きく類別したものである。日本語共起辞書における品詞の種類と、日本語単語辞書における品詞との対応を表7-1に示す。

慣用句フラグは、その形態素と、意味情報（→7.4節）中に示す概念との対応関係の種類を示す情報である。対応関係は、0:（概念が要素番号で指示された形態素に直接対応する）、1:（要素番号で指示された形態素を含む慣用句・複合語に対応する）の2種類である。

概念情報は、その形態素に対応する概念を、概念識別子、日英の概念見出しおよび日英の概念説明で表わしたものである。形態素が非概念語である場合には概念情報は空であり、“ ”と記述される。また、対応する概念が単語辞書中不在概念である場合には日本語による概念説明を補足して記述する。

## 7.3 構文情報

構文情報は、共起句に含まれる表層の共起関係を、共起関係子と受け側・係り側の形態素の3つ組で表した部分構文木である。

日本語共起辞書に記述している共起関係は、格関係、連体修飾関係、連用修飾関係、助数詞による体言の修飾関係の4種類に類別される。格助詞「の」を介した体言間の連体修飾関係や、助詞の「に」「で」などを介した格関係以外の連体修飾関係も格関係に準じて記述されている。

共起関係子には、助詞などの機能語の表記をそのまま用いるものと、用言による連体修飾や連用修飾、助数詞による修飾関係をあらわすための関係子コードを用いるものがある。

助詞相当の機能語などの場合、機能語の表記をそのまま用いる共起関係子は複数の『助詞、助詞相当語、助動詞、助動詞相当語、補助用言』の接続によって構成される。これは、共起項目の内、係り側単語見出しに接続して文節を構成する、いわゆる付属語と考えることができる。日本語共起関係子の種類を表7-2に示す。

受け側・係り側・共起関係子のそれぞれと対応する形態素が、共起句構成要素情報中のどの形態素(列)と対応するかは、それぞれの要素番号列によって示される。ただし、副詞の連用修飾関係のように共起関係が表層の形態素とは対応しないものについては、要素番号列は空となる。

## 7.4 意味情報

意味情報は、句見出しに対応する深層の概念関係を、概念関係子を用いて示した部分意味フレームである。概念関係子は概念辞書で用いられているものと同じである。

意味情報は、句見出しに対応する概念関係を、〈受け側概念要素〉、〈概念関係子〉、〈係り側概念要素〉の3つ組で記述している。〈受け側概念〉と〈係り側概念〉のそれぞれについて、概念が〈共起句構成要素情報〉のどの形態素と対応するかを示す〈受け側/係り側要素番号〉と、概念辞書との対応を示す〈概念識別子〉、〈受け側/係り側単語表記〉が示される。意味情報を持たない共起辞書レコードも存在し、この場合、受け側概念要素、概念関係子、係り側概念要素はすべて“”と記述される。

## 7.5 共起状況情報

共起状況情報は、日本語コーパスにおけるこの共起辞書レコードの出現状況に関する情報を記述したものである。共起状況情報は、共起関係の頻度と共起関係を採った例文の一部からなる。

### 7.5.1 頻度

共起関係の頻度とは、その共起関係が日本語コーパス中で登場した出現回数である。記述される共起頻度は〈表層共起頻度〉、〈共起項目頻度〉、〈共起要素頻度〉の3種類である。

表層共起頻度は、句見出しに対応する表層の共起関係の出現頻度である。共起項目頻度は、構文情報・意味情報に対応する深層の概念関係まで考慮した場合の出現頻度である。共起要素頻度は、共起

句を構成する受け側および係り側の形態素の出現頻度である。共起要素頻度は、表記・かな表記・品詞・概念が同一の要素について集計されている。

## 7.5.2 例文

例文は、日本語コーパスで出現した文を、共起関係に基づき、共起句に含まれる形態素の近傍の形態素を取り出して再構成したものである。例文は、共起関係の成立に影響をあたえる周辺の形態素、例えば否定・アスペクト・使役などに対応する形態素を参照するための情報として使われる。

日本語共起辞書の例文では、受け側である述語に後接する附属語列が共起句を構成する形態素に連結された形式で示される。受け側形態素に対応する部分は'()'で囲まれ、係り側形態素に対応する部分は'<'で囲まれる。係り側と受け側の形態素の間に他の文節がある場合は、「…」を用いて省略を示す。

## 7.6 日本語動詞共起パターン副辞書

日本語動詞共起パターン副辞書は、日本語の主要動詞について、動詞の格に関連する各種の情報を記述したものである。すなわち、各動詞の各概念について、共起しうる表層格の組、表層格に対応する深層格(概念関係子)の種類、および深層格を満たすフィラーとなりうる概念の範囲を記述している。

日本語動詞共起パターン副辞書の主要な役割は、意味解析において、動詞およびそれと共起する名詞に属する複数の概念から、適切な概念を選択するための手がかりを提供することである。

日本語動詞共起パターン副辞書の構造とレコードの例を示す。

===== [日本語動詞共起パターン副辞書レコードの構造] =====

<レコード番号> *1	: レコードタイプと識別番号
<文パターン見出し>	: 共起パターンを文として表現したもの(→7.6.1節)
<構成要素情報>	: 文パターンの構造を表わす構成要素の列(→7.6.2節)
<文パターン構成要素>。。。 *2	: 文パターンを構成する構成要素
<例文>	: 文パターンに適合する例文(→7.6.3節)
<構文情報>	: 共起しうる深層格と表層格の対の集合(→7.6.4節)
<構文情報構成要素>。。。 *3	: 動詞の取りうる深層格と表層格の対
<意味情報>	: 動詞および格のフィラーについての意味情報の集合(→7.6.5節)
<意味情報構成要素>。。。 *4	: 動詞および格のフィラーについての意味情報
<管理情報>	: 辞書開発の管理のための情報
<管理履歴レコード>... *5	: 更新日付等の管理情報

\*1

<レコード番号> ::= JCP <識別番号>

-----

<識別番号> : 7桁の10進数

\*2

<文パターン構成要素> ::= <要素番号> <要素表記> <要素文法情報> <要素意味情報>

〈要素番号〉	：	文中の出現順に従った通し番号
〈要素表記〉	：	文パターンを構成する形態素の表記
〈要素文法情報〉	：	文パターンを構成する形態素の文法情報
〈要素意味情報〉	：	文パターンを構成する形態素の意味情報

## \*3

〈構文情報構成要素〉 ::= 〈概念関係子〉 〈格助詞〉

〈概念関係子〉	：	深層格を表わす概念関係子
〈格助詞〉	：	表層格を表わす格助詞

## \*4

〈意味情報構成要素〉 ::= 〈概念関係子〉 〈概念識別子列〉 〈概念説明列〉

〈概念識別子列〉 ::= 〈概念識別子〉 (；) 。。。

〈概念説明列〉 ::= 〈概念説明〉 (；) 。。。

〈概念関係子〉	：	格を表わす概念関係子
〈概念識別子〉	：	格のフィラーを表わす概念識別子
〈概念説明〉	：	格のフィラーを説明する概念説明

## \*5

〈管理履歴レコード〉 ::= 〈属性名〉 = 〈属性値〉 | 〈管理履歴レコード〉 ; 〈属性名〉 = 〈属性値〉

〈属性名〉 ::= 〈文字列〉

〈属性値〉 ::= 〈文字列〉

=====

```

===== [日本語動詞共起パターン副辞書の例] =====
<レコード番号> JCP0012345
<文パターン見出し> <語1>が<語2>を<語3>にあしらう
<構成要素情報>
<文パターン構成要素>
<要素番号> <要素表記> <要素文法情報> <要素意味情報>
1 <語1>
2 が agent
3 <語2>
4 を object
5 <語3>
6 に goal
7 あしらう 動詞 0e3036 取り合わせる
<例文> /家元/が/松/の/根元/に/菊/を/あしら/う/
<構文情報>
<構文情報構成要素>
<概念関係子> <格助詞>
agent が
object を
goal に
<意味情報>
<意味情報構成要素>
<概念関係子> <概念識別子> <概念説明>
act 0e3036 取り合わせる
agent 30f6b0;30f746 人間;組織
object 30f6ae;444b1a 具体物;具体的あるいは抽象的生産物
goal 30f6ae;444b1a;3aa938 具体物;具体的あるいは抽象的生産物;場所
<管理情報>
<管理履歴レコード> DATE="95/3/31"
=====

```

### 7.6.1 文パターン見出し

日本語動詞共起パターン副辞書の文パターン見出しは、動詞の1つの概念に対する共起パターンを、文として表現したものである。各格のフィラーとなる語(一般には名詞)の位置には、<語1>、<語2>、...、<語n>の形式を持つ語のマークが挿入される。

日本語動詞共起パターン副辞書の格の共起パターンは、いわゆる必須格成分についてのみ記述したものであり、自由格成分は含まない。必須格は日本語単語辞書に登録されている表層格であり、自由格は登録されていない。表層格については、日本語単語辞書の表層格情報についての仕様説明を参照されたい。

### 7.6.2 構成要素情報

日本語動詞共起パターン副辞書の構成要素情報は、前述の文パターン見出しを構成する各要素について、文法情報と意味情報を記述したものである。文法情報としては、構成要素が動詞である場合に、そのマーク(「動詞」)が付与される。意味情報としては、構成要素が動詞である場合には、動詞の語

義を表わす概念の概念識別子と概念説明が付与され、構成要素が助詞である場合には、対応する深層格を表わす概念関係子が付与される。

概念関係子の種類と定義については、概念辞書の概念関係子についての仕様説明を参照されたい。

### 7.6.3 例文

日本語動詞共起パターン副辞書の例文は、前述の文パターン見出しに適合する実例文を表わすものである。文例は形態素分割を行った形式で記述されている。

### 7.6.4 構文情報

日本語動詞共起パターン副辞書の構文情報は、文パターン見出しにおける表層格と深層格の対応を、いわゆる格フレームの形式で記述したものである。構文情報は、深層格を表わす概念関係子と、表層格を表わす格助詞の対の集合からなる。

### 7.6.5 意味情報

日本語動詞共起パターン副辞書の意味情報は、記述対象の動詞と、前述の構文情報に記述された格フレームにおける各格のフィラーとなる語についての意味情報を記述したものである。

動詞については、その語義を表わす概念の概念識別子および概念説明が記述される。

それぞれの格については、格の種類を表わす概念関係子と、格のフィラーとなりうる概念集合の範囲を表わす概念体系上の概念、すなわち、それら概念の集合の共通の上位概念となる概念の集合が記述される。

例えば、「agent 30f6b0;30f746 人間;組織」という意味情報構成要素は、その動詞の概念の agent 格のフィラーとなる概念は、概念体系上の概念「30f6b0 人間」の下位概念か、「30f746 組織」の下位概念のどちらかであると言う判断を示すものである。また、「object 30f6ae-30f6b0 具体物-人間」という意味情報構成要素は、その動詞の概念の object 格のフィラーとなる概念は、概念体系上の概念「30f6ae 具体物」の下位概念のうち「30f6b0 人間」の下位概念でないもののいずれかであるという判断を示す。このように、概念体系上の概念の組み合わせにより、フィラーとなりうる概念の集合を記述する。記号「-」は、指定した概念中の例外となる概念の範囲を示すために使われる。さらに、記号「+」は、例外中の例外(すなわち例外として指定した概念の範囲中で、例外としてフィラーとなりうる概念の範囲)を示すために使われる。

フィラーとなりうる概念体系上の概念は、人間の作業者の判断により記述したものであり、共起辞書または日本語コーパスのデータから抽出したものではない。

### 7.6.6 管理情報

管理情報は、EDR辞書の開発および改良を管理するための情報である。

## 7. a 諸表

表7-1 日本語共起辞書の品詞

表7-2 共起関係子

表7-1 日本語共起辞書の品詞

日本語共起辞書における品詞	日本語単語辞書における品詞	コード
名詞	普通名詞	JN1
	固有名詞	JN2
	数詞	JN3
	時詞	JN4
動詞	動詞	JVE
形容詞	形容詞	JAJ
形容動詞	形容動詞	JAM
副詞	普通副詞	JD1
	陳述副詞	JD2
接尾語	後置助数詞	JN

表7-2 共起関係子

機能語表記(単語列がそのまま関係子となる)

共起関係子	説明・例文(単語表記は、不変化部)
φ	関係子なしで直接修飾できる(名詞、副詞が用言・体言を修飾する場合) [情報, φ, 提供] /<情報>(提供)を開始した。/ [単に, φ, 提供] /<単に>…(提供)するだけでなく/
が	「が」を介した係り受け関係 [会長, が, あいさつ] /<会長>が(あいさつ)した。/
の	「の」を介した係り受け関係 [庭, の, 花] /<庭>の(花)が/
を	「を」を介した係り受け関係 [波紋, を, 投げかけ] /<波紋>を(投げかけ)た。/
に	「に」を介した係り受け関係 [健康, に, い] /<健康>に(い)い/
で	「で」を介した係り受け関係 [自宅, で, 過ご] /<自宅>で(過ご)す/
には	「には」を介した係り受け関係 [寒さ, には, 弱] /<寒さ>には(弱)い。/
への	「への」を介した係り受け関係 [未来, への, 展望] /<未来>への(展望)を/
でも	「でも」を介した係り受け関係 [大学生, でも, 解け] /<大学生>でも(解け)ない/
について	「について」を介した係り受け関係 [弊害, について, 質問] /<弊害>について(質問)する/
:	:
は, へ, から, まで, において, ……	(その他、『助詞、助詞相当語、助動詞、助動詞相当語、補助用言』の連接)

関係子コード

共起関係子	説明・例文(単語表記は、不変化部)
@rentai	用言が連体修飾する(形容詞・形容動詞・動詞が名詞を修飾する場合) [拒否, @rentai, 提案] /<拒否>できない(提案)を/
@renyou	用言が連用修飾する(形容詞又は形容動詞が、動詞又は形容動詞を修飾する場合) [頻繁, @renyou, おこな] /<頻繁>に(おこな)われる/
@unit	名詞と助数詞の組である [台, @unit, 貨車] /5<台>の(貨車)/