

1. はじめに

『日本語歴史コーパス 和歌集編』（以下、『CHJ 和歌集編』）の掛詞情報の付与は、CHJの質的拡張の試みとして、本文の同一箇所に対して複数の形態論情報が付与できるよう開発された「形態論情報の多重化」（小木曾（2016）、村山ほか（2017）を参照）の一つとして行ったものである。以下に『CHJ 和歌集編』の本文に対して行った掛詞情報の付与について記す。

2. 情報を付与した箇所と範囲の認定

『CHJ 和歌集編』の掛詞については、「新日本古典文学大系」（岩波書店）の八代集を収録した分冊『古今和歌集』以下、『後撰和歌集』、『拾遺和歌集』、『後拾遺和歌集』、『金葉和歌集 詞花和歌集』、『千載和歌集』、『新古今和歌集』の脚注解説、神作(2002)『八代集掛詞一覧』に記載のある箇所、範囲とした。

3. 主本文と掛詞（副本文）の区別

コーパス本文における主本文か掛詞かの区別は、国立国語研究所コーパス開発センター（池上尚）編（2016）『『日本語歴史コーパス 平安時代編』形態論情報規程集』の「原則として、掛詞の後ろの語句とのつながりで解釈する。この原則によっても意味を一つに特定できないときは、文脈全体から自然な解釈を選ぶ」という規程に従った。和歌の主たる文脈を作るほうを主本文とし、そうでないほうを掛詞とする。そうしたときに掛詞は、主本文に対して「副本文」と呼ぶこととする。

4. 短単位に付与された掛詞（副本文）情報

掛詞情報の付与にあたっては、UniDicに基づき短単位分割された主本文に対し、掛詞（副本文）も短単位で沿わせている。ただし、一部に次のような対応を行っている箇所があり、検索する上でも注意を要する。

4. 1 掛詞（副本文）情報から除外した例

主本文に対応する掛詞（副本文）がUniDicで同語彙素となってしまうものは掛詞（副本文）としての情報は付与していない。同語彙素か否かの判断は国立国語研究所コーパス開発センター（近藤明日子）編（2016）『近代文語 UniDic 短単位規定集 Ver.1.1』付表を参考とした。以下、例歌と共にサンプルID、開始位置を記しておく。

例) 月のなきには思をきてむねはしりひに心やけをり (20W 古今 0905_21022, 6360)
→下線部分「やけ」について『八代集掛詞一覽』では、景物と心情の両面の意で「焼く」が掛けられているとするが、このような文脈上の意味の区別は UniDic に登録されている語彙素「焼く」自体が担っていないため掛詞 (副本文) 情報は付与しない。

例) あらずのみなりゆくたびの別ぢに (20W 千載 1188_99026,230)
→主本文としては、規程により語彙素「成る」+「行く」と短単位分割されている。そのため、本来掛詞となるはずの「行く」は主本文での「行く」と同語彙素となり、情報は多重とはしない。

4. 2 主・副の関係が複数短単位になる場合

主本文と副本文の関係が2短単位以上にまたがるものについては次のような種類がある。

4. 2. 1 複数短単位にまたがる主本文に対し1短単位の掛詞 (副本文) が対応するもの

例) 足引の山べにをれば白雲のいかにせよとかはる時なき
(20W 古今 0905_10011,13650)
→主本文：語彙素「為る」+「と」+「か」+「晴れる」
副本文：語彙素「ヨドガワ」(地名)

4. 2. 2 1短単位の主本文に対し複数短単位による掛詞 (副本文) が掛かるもの

例) その物ごしのとけずもあらじ (20W 後拾 1087_16017,18590)
→主本文：語彙素「物越し」 副本文：語彙素「裳」+「の」+「腰」

4. 3 掛詞 (副本文) に対する形態論情報付与

掛詞としての副本文に対しても、主本文と同様に形態論情報を付与したが、それらの中には文法上逸脱するものも見られたため、次に例を挙げおく。

4. 3. 1 主本文と掛詞 (副本文) とで未然・連用、終止・連体同形のもの

i. 主本文、掛詞 (副本文) とともに同形の動詞の場合では原則として掛詞 (副本文) の後続語により判断したが、おおむね主本文の活用形に合わせている。

例) たなみのし水ぬるけれど猶かきやらむさてはすむやと
(20W 拾遺 1005_13013,10000)

→主本文：語彙素「住む」動詞-一般/文語四段-マ行/終止形-一般

副本文：語彙素「澄む」動詞-一般/文語四段-マ行/終止形-一般

ii. 後続語による判断ができないものでは連用形、連体形を優先した。

例) しづのまろやのかりの戸を明がたに鳴時鳥哉 (30W 新古 1205_03004,9670)

→主本文：語彙素「明け方」名詞

副本文：語彙素「開ける」動詞-一般/文語下二段-カ行/連用形-一般

4. 3. 2 上記項目には当てはまらないもの

上記の項目に当てはまらないものについては、掛詞（副本文）側で本来有する活用形のいずれかに対応した。この場合、文法的な正しさには反することとなる。

例) 逢事のあけぬ夜ながら明けぬれば (30W 新古 1205_13014,4410)

→主本文：語彙素「明ける」動詞-一般/文語下二段-カ行/未然形

副本文：語彙素「飽く」動詞-一般/文語四段-カ行/已然形-一般

4. 3. 3 掛詞（副本文）が1短単位に満たないもの

主本文に対し、掛かっている語の一部のみが現れており、特定の活用形が想定できないものは活用形の情報付与していない。

例) 涙こそあふみのうみとなりにつれ (20W 後拾 1087_14015,21000)

→主本文：語彙素「オウミ」（地名）

副本文：語彙素「溢れる」動詞-一般/文語下二段-ラ行/活用形は空欄/

出現発音形「オー」

例) 思ふ事おほはら山のすみがまは (30W 新古 1205_17018,20870)

→主本文：語彙素「炭窯」

副本文：語彙素「住む」（動詞-一般/文語四段-マ行/連用形-一般/出現発音形「スミ」）+語彙素「難い」（接尾辞-形容詞的/文語形容詞-ク/活用形は空欄/出現発音形「ガ」）

4. 3. 4 掛詞（副本文）の「出現発音形」

掛詞（副本文）の「出現発音形」は主本文のそれと一致させた。そのため、語として本来は持たない発音形となる場合も生じている。

例) 君にかくあふひや神のしるしなるらん (20W 後拾 1087_19020,7580)

→主本文：語彙素「会う」＋「日」

副本文：語彙素「葵」出現発音形「アウヒ」

例) しら玉をつつむ袖のみながるは (20W 後撰 0955_01001,11260)

→主本文：語彙素「流れる」

副本文：語彙素「泣く」＋「れる」、出現発音形「ナガ／ルル」

4. 3. 5 主本文に対する複数の掛詞（副本文）

参考とした文献において複数の掛詞が記載されている場合、主本文に対し複数の掛詞（副本文）が付くものととらえた。

例) うちつけにこしとや花の色を (20W 古今 0905_10011,7540)

→主本文：語彙素「打ち付け/なり/濃い」

副本文1：語彙素「牽牛子^{ひにごし}」、副本文2：語彙素「実^げに」

5. 掛詞情報が「中納言」検索にヒットする条件指定

上記までに示したようにアノテーションされた掛詞情報は、**そのみを対象として「中納言」による検索・抽出はできず**、以下のような条件指定を行ったときに得られる。

条件指定1

既述したように、和歌集編の掛詞情報は短単位での付与となっている。そのため「中納言」での検索言語単位を短単位とする「**短単位検索**」を選択した時に掛詞情報が得られる。

条件指定2

「中納言」における「検索動作」選択枠の「副本文」のドロップダウンで、「副本文を検索対象に含まない」または「副本文を検索対象に含む」を選択する。（下図は「副本文を検索対象に含む」を選択している状態である。）

検索動作 設定を隠す					
文脈中の区切り記号 ⓘ		文脈中の文区切り記号 ⓘ	#	前後文脈の語数 ⓘ	20
副本文 ⓘ	副本文を検索対象に含む		共起条件の範囲 ⓘ	文境界をまたがない	

和歌集編ではコーパス構築時、可能な限り底本テキストの表層形を優先し、歌の意味解釈において自然なほうを主本文とし、それ以外を掛詞（副本文）としている。「中納言」では副本文が検索対象とされたときに、その検索結果において、掛詞側の形態論情報が表示されるようになっている。

ただし、「検索動作」選択枠における「副本文を検索対象に含まない」を選択した場合、検索自体は主本文のみを対象とした検索となるが、その時、次項目にも挙げる「列の表示」のチェックボックスで「多重化種別」にチェックが入っていれば、検索結果画面の「多重化種別」の列に「掛詞」と表示され、当該箇所に掛詞情報が付与されていることが分かるようになっている。

なお、「副本文を検索対象に含む」を選択した場合、前方・後方共起条件は指定することができないため注意が必要である。

条件指定 3

「列の表示」において、デフォルトのチェックの他に「層」「層内連番」「主本文」「多重化種別」にチェックを入れておく。これらの読み取りについては次節で述べる。

列の表示 ⓘ 設定を隠す

コーパス情報

時代名
 サブコーパス名
 サンプル ID
 開始位置
 連番
 コア
 層
 層内連番
 主本文
 多重化種別

6. 「中納言」での掛詞情報の表示

「中納言」Ver.2.5.0 以降に加わった列表示と、検索結果画面への表示は以下のような内容を意味する。

- (1) 層——該当箇所が有する多重化情報の数。1、2、3重に情報を有する場合、それぞれ10、20、30のように2桁の数で表示される。
- (2) 層内連番——多重化情報の同じ層の中での連番を示す。
- (3) 主本文——検索結果が主本文のものであれば「主本文」列に1を、副本文のものであれば0が表示される。

検索例①：語彙素「恨みる」を「副本文を検索対象に含める」で検索。

- i 検索の結果、主本文の語彙素「見る」（「キー」列）に対する掛詞（「多重化種別」列）としてヒットしている。これは、「主本文」列が「0」となっており、「語彙素」列より右側に示される検索結果が掛詞（「副本文」）としての「恨みる」の形態論情報であることから確認される。
- ii 「層」列の20は当該箇所に対し2層にわたり多重化情報が付与されていることを表している。

サンプル ID	開始位置	層	層内連番	主本文	多重化種別	前文脈	キー	後文脈	語彙素	品詞	活用形
20W古今 0905_15016	23270	20	10	0	掛詞	なれる我が身か # わたつみの我が身にす波立かへりあまのすむてふうら	見	つる哉 # あらをたをあらすき返し返しても人の心をみてこそやまめ	恨みる	動詞-一般	連用形-一般
						夕されは人なきとを打はらひなけかんためとれる我身か # わたつみの我身にす波立かへりあまのすむてふうら	見	つる哉 # あらをたをあらすき返し / \ても人の心をみてこそやまめ			

検索例②：語彙素「腰」を「副本文を検索対象に含める」で検索。

「層」列の 10 は掛詞情報が 1 層付与されていることを表し、同じ層において「腰」が 3 短単位目にあたることを「層内連番」列の 30 が表している。

サンプル ID	開始位置	層	層内連番	主本文	多重化種別	前文脈	キー	後文脈	語彙素	品詞	活用形
20W後拾 1087_16017	18560	10	30		0 掛詞	なん女 のいひをこせて 侍 けれ ば よめる # い こ へ のき なら し 衣 いまさら に そ の	物 こ し	の と け ず し も あ ら し # 大式 資 通 む つ ま し き さ ま に な ん い ふ と き き て つ か は し け る # まこと に	腰	名詞-普通名詞-一般	
						たりける返事 に、ものこ し に なん女 のいひをこせて侍 けれ ば よめる # い こ へ のき なら し 衣 いまさら に そ の	物 こ し	の と け ず し も あ ら し # 大式 資 通 む つ ま し き さ ま に な ん い ふ と き て つ か は し け る # まこと に や 空 に な き 名 の ふ り ぬ ら ん あ ま て る 神のくもりなきよに			

参考文献

- 小木曾智信（2016）「多重の読みを持つテキストのコーパス化」『言語資源活用ワークショップ 2016 発表論文集』国立国語研究所（<http://doi.org/10.15084/00001470> にて公開）
- 神作光一編（2002）『八代集掛詞一覧』風間書房
- 国立国語研究所コーパス開発センター（池上尚）編（2016）『『日本語歴史コーパス 平安時代編』形態論情報規程集』（https://pj.ninjal.ac.jp/corpus_center/chj/doc/morph-heian-2016.pdf にて公開）
- 国立国語研究所コーパス開発センター（近藤明日子）編（2016）『近代文語 UniDic 短単位規定集 Ver.1.1』（https://pj.ninjal.ac.jp/corpus_center/chj/doc/unidic-MLJ_rulebook_v1_1.pdf にて公開）
- 松崎安子・小木曾智信・中村壮範（2019）『『日本語歴史コーパス 和歌集編』Ver.1.0 の公開』『日本語学会 2019 秋季大会予稿集』
- 村山実和子・小木曾智信・中村壮範（2017）「形態論情報の多重化による洒落本コーパスの質的拡張」『研究報告人文科学とコンピュータ（CH）』2017(8)（<http://id.nii.ac.jp/1001/00178697/> にて公開）