

『現代日本語書き言葉均衡コーパス』データに基づく字音接辞の共起成分データベースの構築

秦 周漢 (総合研究大学院大学)

Construction of a Database for Co-occurring Components of Sino-Japanese Affixes Based on the BCCWJ Corpus

Zhouhan Qin (The Graduate University for Advanced Studies)

要旨

日本語の漢語において、造語力の強さはその顕著な特徴の一つである。特に字音接辞は漢語の語形成に重要な役割を果たしているが、従来の研究では、接辞の使用実態に関する計量的な特徴については十分に明らかにされていない。本研究では、張 (2019) の日本語字音接辞語彙表を参考にし、『現代日本語書き言葉均衡コーパス』に含まれる字音接辞および、その接辞と共起する語基成分を抽出し、手作業で語彙情報の修正を行い、字音接辞の共起成分データベースを構築した。各接辞とその共起成分の頻度のほか、『分類語彙表 増補改訂版』に基づく意味分類情報も付与した。本データベースは、共起成分の頻度情報を付与することにより、共通の意味成分の抽出や生産性の数値化に道を拓くものであり、中国語接辞との対照研究も可能にする。習得研究や日本語教育や国語教育への貢献も期待される。

1. はじめに

一般的には、接辞とは、単独で使用することはできず、他の語基と結合して派生語を形成することができる拘束形態素を指す。この定義を日本語に適用すると、和語系の形態素には、概ね当てはまる。例えば、「大人っぽい」の「-っぽい」、「高み」の「-み」などの接辞成分がある。しかし、この定義を字音形態素に適用すると、やや不適切なケースがある。水野 (1987) は、「営業部」の「-部」や「大規模」や「大-」など、一見すると接辞に見える語が「部がちがう」、「(サイズの) 大はありません」という単独の使い方になる例を挙げている。これは接辞の定義に反する。したがって、研究者の立場によって、字音接辞の捉え方も異なる。

最も狭い定義は森岡 (1994) の記述に見られる。森岡 (1994) は、漢語系の形態素には接辞がないか、あるいは「御 (ご)」のみが接辞と言えると主張し、「営業部」の「-部」や「大規模」の「大-」などの語を「準接辞」と呼んでいる。前述した水野 (1987) は、「結合形式であることを、接辞の条件としてあまりきびしく考えないほうが良いのではないか」という緩い主張を持っている。

本稿は、同様に緩い主張を持っている山下 (2013) の接辞観を支持する。

字音形態素が造語 (語形成) の上で、語基的にも接辞的にも振る舞うことがあるという事実を捉え、字音形態素が造語 (語形成) において果たす機能によって、それを語基と見なすか、接辞と見なすかが決まると考える。つまり、一つの字音形態素は語基か接辞かどちらに分類されるものではなく、我々認知主体が、個々の言語表現における語について、造語機能のどこを焦点化して認知するかによって、字音形態素は語基にも接辞にもなり得ると捉えるわけである。

(山下 2013 : 68)

静的な見方と比べて、この主張は動的な立場から形態素を捉えている。すなわち、「接辞」を定義するのではなく、形態素の性質や用法を描写する語であると考えている。したがって、山下（2013,2018）の論述では、「非」「大」「亜」「者」「化」などの接辞的な成分について、それを接辞とは呼ばず、「接辞性字音形態素」と呼んでいる。本稿では、これらの接辞となり得る「接辞性字音形態素」を、便宜上、以下「字音接辞」と呼ぶ。

2. 字音接辞に関する包括的な計量研究

今までの字音接辞研究は、一つまたは複数の特定の接辞に焦点を当て、詳細に分析するのがほとんどで、接辞全般を対象に量的な特徴を示す研究は多くないが、ここでは代表的な研究として、野村（1978）および山下（2008,2013,2018）を取り上げる。

まず、野村（1978）では、接辞的に使用される語を「接辞性字音語基」と呼び、以下のよう

に定義した。

すでに存在する、和語・外来語の語基、および、字音複合語基、そして、それらの結合形に、前部分あるいは後部分から結合する、字音形態素。

（野村 1978：104）

そして、野村は国立国語研究所（1976）の『現代新聞の漢字』という資料を主なデータとし、接辞性字音語基全般を対象に、考察を行っている。前部分の接辞性字音語基を 250 語、後部分の接辞性字音語基を 605 語抽出し、その使用頻度上の特徴と結合力について詳細に論じた。その後、品詞性、前後の語の結合関係、意味分類の観点から各接辞を以下の表 1 のように分類した。

表 1. 野村（1978）の接辞の分類

接頭辞		接尾辞	
分類	例	分類	例
体言型	党大会・核爆発	体言型	先進国・日本人
連体修飾型	大都市・中学校	用言型	近海産・外国製
連用修飾型	再検討・最年少	相言型	積極的・柔軟性
連体詞型	同議員・前会長		
用言型	反政府・超党派		
否定辞型	無意識・不明朗		
数量限定型	第○日・約○分		
敬意添加型	御婚礼・令夫人		

次に、山下（2008）は、まず、国語辞典を中心とした資料から、接辞と造語成分を抽出して、データ数約 4,800 の『造語成分データベース』を作成した。「見出し語・標準表記・接続・語種・語例・9 種資料出現頻度・各資料収録状況」などの情報を付与している。次に『造語成分データベース』を基にして、これらの造語成分を含む合成語を『分類語彙表増補版』

および『日本語能力試験出題基準語彙表』から抽出し、およそ 32,600 語の合成語を得た。それを基に『合成語データベース』を作成した。このデータベースには「見出し語・よみ・品詞・語種・造語成分・結合語基・新聞出現頻度・Web 出現頻度」などの情報を付与している。

しかし、山下 (2008) のデータベースは接辞を対象に構築されたものではなく、そのデータベースには接辞のほか、接辞ではない造語成分も含まれている。したがって、山下 (2008) のデータベースを利用して接辞を分析するためには、接辞に特化したスクリーニングが必要である。そのため、山下 (2013) の研究では、山下 (2008) のデータベースから接辞を選別し、合計で接頭辞 169 個、接尾辞 589 個、接頭接尾辞 80 個を選出した。その後、接辞の造語機能と分類を検討した。接辞の造語機能について、「結合機能・意味添加機能・品詞決定機能・文法化機能」という四つの点があるとしている。

山下 (2013) の研究では、選出された接辞の語彙表は提示されていないため、そこで補足として、山下 (2018) ではより詳細な考察が行われている。山下 (2018) の研究では、山下 (2013) の研究を基にして、接辞の数も更新されており、接頭辞は 234 個、接尾辞は 656 個となっている。接辞の分類については、野村 (1978) と違い、意味の観点から分類されており、最新の山下 (2018) の研究では表 2 のような分類がなされている。

表 2. 山下 (2018) の接辞の分類

接頭辞		接尾辞	
分類	例	分類	例
形容	亜熱帯・正社員	事物	作業衣・換気扇
指定	各学校・現会長	類別	第一位・十歳
事物	核実験・乳製品	空間範囲	支配下・富士山
行為作用	過保護・誤発射	精神	修正案・経済学
強調	最先端・小都市	人物	事務員・警察官
精神	気高い・税負担	行為作用	予備役・温暖化
待遇	令夫人・御案内	組織集団	会計課・図書館
可否	非常識・不景気	時間	営業中・六十秒
空間範囲	駅職員・内出血	数量程度	日本一・最低限
組織集団	県知事・市役所	様相	連体形・回転式
		分別	芸能界・製造業
		待遇	山田君・今川氏

上記の研究において、野村 (1978) の研究は国立国語研究所 (1976) の調査資料を使用している。山下 (2008) が構築したデータベースは、接辞に特化したものではなく、「広がる」の「-がる」のような接辞以外の造語成分が含まれている。そして、このデータベースの合成語は『分類語彙表増補版』と『日本語能力試験出題基準語彙表』から選ばれたものであり、常用の合成語しか含まれていない。例えば、接尾辞「式」について、山下 (2008) の『合成語データベース』には 28 個の合成語しか含まれていない。「公式」「形式」などの二字漢語の

複合語を除く、「結婚式」などの接辞の用法はわずか8例しかない。生産性のような接辞の造語機能を分析する場合、辞書や語彙表ではなく、コーパスから抽出された派生語を用いる必要がある。コーパスのほうがより实际的で広範な言語使用状況を提供できるからである。辞書や語彙表には通常、規範的な語彙しか含まれておらず、多様性に欠ける。

したがって、日本語における字音接辞を全体的に詳細に分析するためには、大規模コーパスを基づく字音接辞データベースを構築する必要がある。

3. 字音接辞の選定：張（2019）の研究

本研究で構築したデータベースに使用される字音接辞は、張（2019）の研究から選ばれたものである。張（2019）は、山下（2013）と同様に動的な立場で、前述の野村（1978）の「接辞性字音語基」の定義を基に、常用漢字表の2,136字を対象に、7冊の国語辞典の用例を参考にし、調査を行った。二字（以上）漢語や、和語、外来語と結合する「用例」が国語辞典で提示されている場合に接辞と認める。例えば、「式」を例にとると、

- (1). 式を挙げる 式で表わす
- (2). 卒業式 スパルタ式 西洋式 電動式 方程式
- (3). 形式 旧式 公式

（張 2019 : 8）

以上の三つの用法について、張（2019）の研究では、(1)のような使い方を「語基」、(2)のような使い方を「接辞」、(3)のような使い方を「二字漢語の構成要素」と呼んでいる。調査結果としての語彙表には、接頭辞として使用できる漢字は223字、その中で音読みが二つある漢字は7字ある。接尾辞として使用できる漢字は546字、その中で音読みが二つある漢字は16字ある。また、これらのうち、接頭のかつ接尾的な漢字は76字存在する。

4. データの検索および抽出

本研究のデータは『現代日本語書き言葉均衡コーパス』（以下、BCCWJ）から取得され、検索エンジン「中納言」を利用して検索した。中納言のバージョン情報は「中納言 2.7.2 データバージョン 2021.03」である。

検索方法としては、短単位検索の条件式検索を使用し、各字音接辞を共起成分として指定し、かつ共起成分（接辞）の語種を「漢語」にして、条件を指定しない「キー」で検索を行った。例えば、「御」を接頭辞とする派生語および、「者」を接尾辞とする派生語の検索条件式は以下の通りである。

接頭検索式：前方共起: (語彙素="御" AND 語種="漢") ON 1 WORDS FROM キー

```
WITH OPTIONS tglKugiri="|" AND tglBunKugiri="#" AND limitToSelfSentence="1" AND  
tglFixVariable="2" AND tglWords="10" AND unit="1" AND encoding="UTF-16LE" AND  
endOfLine="CRLF"
```

接尾検索式：後方共起: (語彙素="者" AND 語種="漢") ON 1 WORDS FROM キー

```
WITH OPTIONS tglKugiri="|" AND tglBunKugiri="#" AND limitToSelfSentence="1" AND
```

```
tglFixVariable="2" AND tglWords="10" AND unit="1" AND encoding="UTF-16LE" AND endOfLine="CRLF"
```

以上の検索式によって、検索されたキーには、「(御) 苦勞」「(御) 主人」「勞働 (者)」「消費 (者)」などの語基だけでなく、以下のようなノイズも含まれる。

(4). 連署であられた叔父御なら政所のすべての者が得心いたすはず

(BCCWJ サンプル ID : PB19_00613 高橋克彦 (著) 『時宗』)

(5). 松前藩は、桜井を、「裏切り者」として処刑した。

(BCCWJ サンプル ID : PB59_00730 富樫倫太郎 (著) 『美姫血戦』)

(4)での「御」は「叔父御」という派生語の接尾辞として存在するが、接頭辞の検索式で検索すると誤って検出されている。このような検索の錯誤は、接頭辞・接尾辞の両方で使用できる76の漢字において大量に存在する。(5)での「裏切り者」について、検索時に接辞の語種を漢語に指定しているにもかかわらず、BCCWJの誤解析により、このようなケースが時々発生する。これらの誤検出は、人手によるチェック時に削除する。

ここで特に述べておくべき点がある。検索時に、条件を「品詞＝接尾辞」あるいは「品詞＝接頭辞」を追加して指定することで、その漢字が接辞として使用される際の共起成分を特定して検索することができる。しかし、本研究ではそのような検索方法を取っていない。その理由は以下の二点である。

第一に、BCCWJの短単位語彙表 (Version 1.0) を確認すると、張 (2019) の研究で示されている前接辞用法の223字のうち、実際に接頭辞としてリストされているのは116字のみで、後接辞用法の546字のうち、実際に接尾辞としてリストされているのは315字しかない。すなわち、検索時に品詞を指定することで本来取り上げるべき接辞を検索できなくなる可能性が高いのである。

第二に、BCCWJには自動解析の誤差が存在するため、接辞用法が正しく接辞として解析されない場合がある。例えば、「結婚式」という派生語はBCCWJでの出現頻度が1,924回であり、その中で「-式」の品詞が「名詞-普通名詞-一般」として解析されるケースは1,862回、「接尾辞-名詞的-一般」として解析されるケースはわずか62回である。

これらの不安定な要因を考慮し、品詞の指定なしで検索を行うことにした。また、「第一、年、者、人、的、様」の7字の検索結果は10万項目を超えてしまい、年代別に検索してまとめた。

上述の検索を経て、各接辞と共起する語の情報が含まれるファイルを得た。その後、各ファイルの「語彙素」列をもとに各接辞の共起成分の異なり語数を集計し、まとめた結果、合計402,930項目が含まれていた。76,353項目の固有名詞と数詞を除き、人手によるチェックを経た現段階のデータベースには、字音接辞を含む派生語が合計175,930項目含まれている。張 (2019) の基準に従って、共起成分はすべて二字 (以上) 漢語や、和語、外来語である。一字漢字は含まれていない。

5. 各見出し項目の情報

データベースの各見出し項目は表3に示す通りである。以下、派生語および接辞の量的特徴について簡単に示す。

表 3. データベースの各見出し項目

番号	見出し項目	説明
1	派生語	語基と接辞が結合して形成する新しい語
2	接続型	接頭的用法か、接尾辞的用法か
3	接辞	張 (2019) による接辞として使われる漢字
4	接辞読み	接辞として使われる漢字の読み方
5	語基	接辞と共起する語基
6	語基読み	接辞と共起する語基の読み方
7	語基分類	『分類語彙表増補改訂版』による語基の意味分類
8	語基語種	BCCWJ から抽出する語基の語種情報
9	語基品詞	BCCWJ から抽出する語基の品詞情報
10	派生語頻度	BCCWJ 全体における派生語の使用頻度
11	出版・新聞	出版・新聞における派生語の使用頻度
12	出版・書籍	出版・書籍における派生語の使用頻度
13	出版・雑誌	出版・雑誌における派生語の使用頻度
14	図書館・書籍	図書館・書籍における派生語の使用頻度
15	特定目的・ベストセラー	特定目的・ベストセラーにおける派生語の使用頻度
16	特定目的・広報紙	特定目的・広報紙における派生語の使用頻度
17	特定目的・教科書	特定目的・教科書における派生語の使用頻度
18	特定目的・知恵袋	特定目的・知恵袋における派生語の使用頻度
19	特定目的・韻文	特定目的・韻文における派生語の使用頻度
20	特定目的・ブログ	特定目的・ブログにおける派生語の使用頻度
21	特定目的・国会会議録	特定目的・国会会議録における派生語の使用頻度
22	特定目的・白書	特定目的・白書における派生語の使用頻度
23	特定目的・法律	特定目的・法律における派生語の使用頻度
24~60	1971~2008_出版年	各出版年における派生語の使用頻度

5.1 派生語の出現頻度

派生語は、検索された共起語基と接辞の組み合わせによって構成される。検索時に「接辞」を指定する条件を追加しなかったため、「いえ別」「多い様」「の他」「金詐欺」など、多くのノイズも検索結果に含まれている。これらのノイズを削除した後の 175,930 項目の中には、接頭辞派生語が 32,584 項目、接尾辞派生語が 143,346 項目含まれている。BCCWJ (総語数約 1 億語) におけるすべての派生語の出現頻度の分布状況は表 4 に示す通り、頻度の多くは 0~1000 の範囲に集中していることがわかる。

表 4. BCCWJ における派生語の頻度分布と派生語例

頻度区間	異なり語数	異なり語数全体に占める比率	派生語例	語例の頻度
0 - 1,000	175,640	99.835%	競争力	962
1,000 - 2,000	184	0.105%	高齢化	1,975
2,000 - 3,000	49	0.028%	政治家	2,969
3,000 - 4,000	28	0.016%	不可能	3,908
4,000 - 5,000	12	0.007%	弁護士	4,851
5,000 - 6,000	6	0.003%	事務所	5,929
6,000 - 7,000	4	0.002%	労働者	6,980
7,000 - 8,000	2	0.001%	積極的	7,583
8,000 - 9,000	0	0.000%	/	/
9,000 - 10,000	1	0.001%	基本的	9,965
10,000 - 11,000	2	0.001%	具体的	10,550
11,000 - 12,000	0	0.000%	/	/
12,000 - 13,000	0	0.000%	/	/
13,000 - 14,000	0	0.000%	/	/
14,000 - 15,000	1	0.001%	可能性	14,502
15,000 - 16,000	1	0.001%	委員会	15,802

1～1000 の頻度区間の分布状況を描いた頻度分布図は図 1 の通りである。全体としてはベキ分布を示しているが、多くの派生語は低頻度の区間に集中している。計算の結果、1～40 の頻度区間での派生語が全体の派生語の 94.7%を占めていることが分かった。その部分の拡大図も図 1 に示す。また、1～10 の頻度区間の派生語数と全体派生語で占める比率は表 5 に示した。

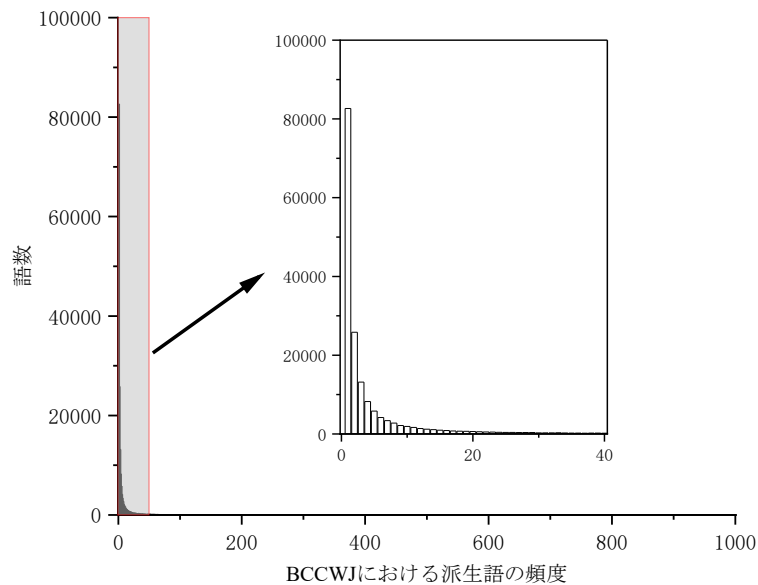


図 1. BCCWJ における派生語の頻度区間 1～1000 の語数分布

表 5. 1～10 頻度区間の派生語の比率と語例

頻度	異なり語数	異なり語数全体に占める比率	派生語例
1	82,649	46.978%	固有化
2	25,837	14.686%	耕作人
3	13,175	7.489%	青年派
4	8,247	4.688%	主業務
5	5,816	3.306%	忠義心
6	4,171	2.371%	航空士
7	3,383	1.923%	伸縮式
8	2,757	1.567%	不安全
9	2,157	1.226%	除湿剤
10	1,917	1.090%	整備用

以上の特徴から、接辞と語基の組み合わせによる派生語の大部分は、慣用度が低く、辞書に収録されない語であることがわかった。BCCWJにおいて出現頻度が1の派生語は、全体の約半分に達しており、これは接辞の生産的な機能を反映している。つまり、日本語の語彙の拡張において字音接辞は非常に重要な役割を果たしている。前述したように、山下(2008)のデータベースにおいて、接辞としての「式」の派生語は8例しかなく、これらは辞書や単語表から抽出されたもので、常用の語彙を代表している。一方、本データベースには、接辞としての「式」の派生語は合計で1,373語も収録されている。

ここでは、例として、出現頻度が最も高い接頭辞派生語と接尾辞派生語を上位20語ずつ表6に示す。

表 6. 出現頻度上位20の派生語

接頭		接尾	
小学校	6,904	委員会	15,802
小企業	4,593	可能性	14,502
不動産	3,911	具体的	10,550
不可能	3,908	自動車	10,273
中学校	3,621	基本的	9,965
御指摘	3,334	積極的	7,583
大規模	3,130	高齢者	7,491
前年度	3,072	労働者	6,980
御存知	2,725	一般的	6,394
市役所	2,665	消費者	6,236
被保険	2,230	社会的	5,987
新幹線	2,158	事務所	5,929
大部分	2,031	障害者	5,795
大都市	1,929	裁判所	5,470
御主人	1,917	図書館	5,326

御意見	1,809		事業者	5,292
不安定	1,776		自転車	4,911
県知事	1,735		弁護士	4,851
大企業	1,547		保険料	4,745
無意識	1,515		外国人	4,660

5.2 異なり派生語数および延べ派生語数

上述のように、本データベースには合計 175,930 項目のがあり、その中には接頭辞派生語が 32,584 項目、接尾辞派生語が 143,346 項目含まれている。BCCWJ における延べ語数は総計 2,679,470 語、延べ接頭辞派生語数は 422,788 語、延べ接尾辞派生語数は 2,256,682 語である。BCCWJ の総語数は短単位で合計 104,911,464 語、長単位で合計 83,585,665 語であるため、本データベースの接辞は短単位の約 2.6%、派生語は長単位の約 3.2%を占めているということになる。

異なり派生語数と延べ派生語数が最も高い接頭辞および接尾辞をそれぞれ 20 語ずつ表 7 に示す。

表 7. 出現頻度上位 20 の接辞

延べ派生語数				異なり派生語数			
接頭		接尾		接頭		接尾	
御	57,367	的	265,182	大	2,698	等	7,203
大	37,938	者	180,677	各	2,283	的	5,358
不	32,952	性	74,334	御	2,190	用	3,873
小	19,354	化	65,552	新	1,760	中	3,388
各	16,536	等	62,899	全	1,384	者	2,839
新	16,261	会	52,458	同	1,280	性	2,830
無	14,612	法	36,936	非	992	化	2,733
再	12,710	所	36,799	超	874	後	2,269
非	10,211	中	34,901	両	862	法	2,258
同	9,025	人	34,897	小	855	系	2,160
高	8,522	力	27,456	旧	827	時	2,044
総	7,688	書	25,685	無	745	上	1,966
全	7,181	用	23,547	再	724	部	1,923
諸	7,096	上	23,456	総	656	内	1,738
副	5,584	部	23,428	諸	644	論	1,389
一	5,521	車	23,269	不	550	数	1,386
前	5,454	館	23,106	対	541	式	1,363
超	5,027	後	22,916	高	516	風	1,240
未	4,933	省	22,840	本	458	分	1,145
低	4,908	家	22,788	兼	437	人	1,111

6. おわりに

先行研究から、日本語の接辞についてはまだ多くの議論の余地があることがわかる。したがって、本研究では接辞に特化したデータベースを設計、開発した。しかし、本データベースにはいくつかの限界がある。張 (2019) の研究は常用漢字表から抽出された漢字を使用しているため、数の上では野村 (1978) や山下 (2013) の研究に比べて少なく、日本語全体の字音接辞を網羅しているとは言い切れない。さらに、本データベースの共起語基はすべて二字 (以上) 漢語であり、そのため、派生語もすべて三字 (以上) 漢語となっている。二字漢語は議論の対象外とした。また、野村 (1987) の研究では、「国際」のような二字接辞¹と見なせる語彙素にも言及されているが、これも本データベースには含まれていない。今後、本データベースはさらに改善を続け、これらの問題を順次解決していく予定である。また、字音接辞に限らず、和語系接辞や外来語系接辞も将来的にはデータベースに追加する予定である。

本接辞データベースは、日本語の語彙研究にとどまらず、日本語教育にも応用可能である。字音接辞は日本語において非常に多く、強力な派生能力を持つため、学習者にとって接辞の意味とその派生用法を習得することは重要である。日本語学習の観点から見ても、本データベースは日本語接辞を詳細に理解し、研究したい教師、学習者、研究者にとって有益なツールとなるはずである。さらに、筆者の研究の一環として、今後は本データベースを基に、中国語接辞の対応情報も追加し、日本語と中国語の対照研究も行いたい。また、データベースをもとにサンプリングを行い、教育現場で利用できる接辞テストを開発し、接辞の習得の研究を行いたいと考えている。

付 記

本データベースの名称は、暫定的に『日本語派生語データベース (字音接辞版)』とする。以下の URL からダウンロードできる。

URL: <https://github.com/zhouhan-QIN/DB-for-JDW>

文 献

- 阪倉篤義 (1986). 「接辞とは」 『日本語学』, 5:3, pp.4-10.
- 張 明 (2019). 『現代日本語における字音接辞の研究-連体詞型字音接頭辞の記述的研究を中心に』 学習院大学, 博士論文.
- 野村雅昭 (1977). 「造語法」 『岩波講座日本語 9 語彙と意味』, pp.245-284.
- 野村雅昭 (1978). 「接辞性字音語基の性格」 『国立国語研究所報告 61 電子計算機による国語研究IX』, pp.102-138.
- 野村雅昭 (1987). 「複合漢語の構造」 『朝倉日本語新講座 1 文字・表記と語構成』, pp.130-144.
- 野村雅昭 (1988). 「二字漢語の構造」 『日本語学』, 7:5, pp.44-55.
- 野村雅昭 (1999). 「字音形態素考」 『国語と国文学』, 76:5, pp.1-10.
- 野村雅昭 (2013). 「品詞性による字音複合語基の分類」 野村雅昭(編) 『現代日本漢語の探究』, pp.134-145.
- 水野義道 (1987). 「漢語系接辞の機能」 『日本語学』, 5:3, pp.60-69.

¹ 野村 (1987) では、「結合専用語基」と呼ぶ。

- 森岡健二 (1994). 『日本文法体系論』 明治書院.
- 山下喜代 (2008). 『日本語教育のための合成語データベース構築とその分析』 (平成 17 年度～平成 19 年度科学研究費補助金研究成果報告書).
- 山下喜代 (2013). 「接辞性字音形態素の造語機能」 野村雅昭(編)『現代日本漢語の探究』, pp.83-108.
- 山下喜代 (2018). 「字音形態素のカテゴリー化：接辞を中心に」 『青山語文』, 48, pp.217-228.

関連資料

- 国立国語研究所 (編) (1976). 『現代新聞の漢字』 秀英出版.
- 国立国語研究所 (編) (2004). 『分類語彙表 増補改訂版』 大日本図書.
- 国立国語研究所 (2011). 『現代日本語書き言葉均衡コーパス』 (データバージョン 2021.03, 中納言バージョン 2.7.2) (2023 年 5 月 12 日確認).

関連 URL

- | | |
|---------------------|---|
| コーパス検索アプリケーション『中納言』 | https://chunagon.ninjal.ac.jp/ |
| 分類語彙表－増補改訂版データベース | https://github.com/masayu-a/WLSP |