

計算機用日本語基本名詞辞書における 見出し語の下位区分

桑畑和佳子, 橋本三奈子

情報処理振興事業協会(IPA) 技術センター

{kuwahata, hasimoto}@stc.ipa.go.jp

【IPAL名詞辞書】では見出し語のもつ用法を意味からだけではなく統語的特徴からも分類し、これを「下位区分」と呼んでいる。名詞辞書の記述を進める際はこの「下位区分」を一貫して行うことが重要である。そこで、「下位区分」の客観的な判断基準として、形式的な基準と意味的な基準と、あわせて11通りの基準を設けた。これらは相互に関係しあっていて、どれか一つの基準だけで十分であるということは少なく、いくつかの基準を組み合わせて「下位区分」するものである。本稿では、それらの基準を組合せて用いて、総合的な判断を行いながら「下位区分」をした具体例をあげ、名詞辞書における「下位区分」の方法について述べる。

On the methodology of subdivision for the Lexicon of Basic Japanese Nouns

KUWAHATA Wakako, and HASIMOTO Minako

Software Technology Center,

INFORMATION-TECHNOLOGY PROMOTION AGENCY, JAPAN

(Shuwa-Shibakoen 3-chome BLDG 3-1-38 Shibakoen, Minato-ku, Tokyo, 105 JAPAN)

This paper proposes eleven criteria for the subdivision of nouns in the Lexicon of Basic Japanese Nouns. Some of the criteria are based on grammatical information. The others are based on semantical information. Using only one criterion is not enough to subdivide the noun in terms of its semantical features. In this paper, we shall combine these criteria, showing some examples in each subdivision by means of our comprehensive judgment. We shall also focus on the effective division of using these eleven criteria.

1. はじめに

I PA技術センターではこれまでに動詞と形容詞の計算機用辞書を作成し『計算機用日本語基本動詞辞書 I P A L (Basic Verbs)』『計算機用日本語基本形容詞辞書 I P A L (Basic Adjectives)』として公開してきた。I P A Lの特徴は、機械翻訳、ワープロ仮名漢字変換など、どのようなアプリケーションにおいても、日本語を処理するにあたって必要になる情報がふんだんに盛り込まれているところにある。現在開発中の『計算機用日本語基本名詞辞書 I P A L (Basic Nouns)』（第一版をftp公開中）も、名詞のもつ構文内部の働きに焦点をあてた点で、『動詞辞書』と『形容詞辞書』の流れをひきつぐものといえる。I P A Lでは、見出し語のもつ用法を意味だけからではなく統語的特徴からも分類し、これを「下位区分」と呼んでいる。本論文では『名詞辞書』における「下位区分」の方法について述べる。

2. 「慣用句」と下位区分

『名詞辞書』では下位区分ごとに、

- A. 形態情報
- B. 意味情報
- C. 述語の項としての用法
- D. 連体修飾語としての用法
- E. 連体被修飾語としての用法
- F. 述語としての用法
- G. サ変動詞としての用法

を記載した[1]。

この他、下位区分はしないが、見出し語の用例として記載したものとして、「慣用句」がある。「慣用句」は、(1)個々の語の意味の総和が当該句全体の意味にならない、(2)述語の活用形や使役化、否定化、丁寧化や、さらに修飾語の付加や省略の可能性などに制限が見られる、という特徴を持つ。例えば、次のものがそうである。

- 1) 鼻が高い

この表現が「得意になっている」という意味で用いられている時は、「鼻」には実際の「鼻」の意味がない。このような場合を「慣用句」と呼ぶ。『名詞辞書』では、一つの見出し語につき一つの「慣用句」欄を設け、その欄にまとめて記載してある。

一方、「鼻が大きい」「鼻がいい」の場合には、それぞれ、「動物の顔にあり、呼吸や発声を助ける器官」「鼻」の機能・働きのように、句全体の意味を個々の語の意味へ還元することができ、文法的なふるまいに制限がない結びつきである（これを「コロケーション」と呼ぶ[2]）。そこで、これらは見出し語の一般的な用法であると認め、それぞれ一つの下位区分をたて、先にあげたAからGまでの情報の記述対象とした。

中には、「慣用句」とすべきなのか、見出し語の一つの「コロケーション」として下位区分をたてて記述すべきなのか、判断に迷うものがある。次にあげる見出し語「腕」の例は、そのようなものの一つである。

2) 腕がいい

身体部位としての「腕」を指してはいないので、これは「慣用句」になるのであろうか。この時の「腕」には次のように複数の「コロケーション」がある。

- 3) 腕が確かだ／上がる／鈍る／立つ
- 4) 腕を磨く／上げる／競う／ふるう

これらの結びつきから、「腕」に「技量」という意味を認めて、下位区分を一つ設け、3)、4)は「述語の項としての用法」として記述すべきであると判断できる。

「腕」の例とは違い、複数のコロケーションを持たず、結びつく述語が三つより少ないものについては、「慣用句」とした。

次の見出し語「虫」の例では、「第六感」という意味を認めることもできそうであるが、結びつく語の数が限られているため、「慣用句」と判断した。

- 5) 虫が知らせる
- 6) 虫の知らせ

3. 「下位区分」の判断基準

辞書記述の際には、「慣用句」ではないという判断をした後に、複数ある用法を「下位区分」するが、実はこの判断は、「慣用句」であるか否かの判断よりも複雑である。結びつく述語の違いが顕著であれば、そこに見出し語の用法の違いを見出しやすいが、実際には述語にくる動詞や形容詞なども多義であるため、見出し語の用法が違って結びつく述語に差はなく、それだけでは「下位区分」の判断がつかないものが多い。また、辞書を共同作業で執筆する場合、人によって下位区分数が多かったり少なかったりという問題も生じる。そこで、「下位区分」する際の基準をほかにも幾つか設け、判断の助けとした。

「下位区分」の判断基準として用いるものを列挙すると、以下の通りである。（「NP0」は見出し語を表す）

A 形式的な基準

1. 結びつく述語の違い
2. サ変動詞用法や述語用法の有無の違い
3. 連体修飾語の違い（NPの／なNP0）
4. 連体被修飾語の違い（NP0の／なNP）
5. 合成語の違い

B 意味的な基準

6. 意味カテゴリの違い
7. 意味素性の組み合わせの違い
8. 類義語や対語の組み合わせの違い
9. 比喩的表現か否かの違い
10. 総称か特化かの違い
11. 対訳の違い

これらは相互に関係しあっていて、どれか一つの基準だけで十分であるということは少なく、いくつかの基準を組み合わせる「下位区分」するものである。

4. 形式的な基準による下位区分

4.1 「結びつく述語の違い」を用いた例

最初に、「結びつく述語の違い」が顕著に見られる見出し語「根」の例を示す。

- 1) 根 がつく／が伸びる／が生える／が枯れる
- 2) 根 が深い／を絶やす／を絶つ／をおろす
- 3) 根 が優しい／が明るい／が卑しい

上で見られる述語の違い通り、次のように下位区分される。

1. 植物を支え、水分／養分を吸収するため、
地中にある器管。 1)
2. 抽象的な物事のもととなっている部分。 2)
3. 生まれ持った性格。 3)

しかし、このように一つの基準だけで分けられる名詞は実際には少ない。「結びつく述語の違い」だけでなく、様々な情報を用いての総合的な判断が必要なことは、以下に述べる通りである。

4.2 「サ変動詞用法の有無」を用いた例

見出し語「収穫」には次のような用例がある。

- 1) ブドウの収穫 が 始まる／終わる
- 2) 大豆の収穫 が ある／ない／
多い／少ない／
減る／増える
- 3) 今日の会議は収穫 が ある／ない／
多い／少ない

2)と3)とでは、「減る、増える」を除いて、「ある、ない、多い、少ない」など、結びつく述語に類似性がみられるが、ここから2)は1)よりも3)に近いと考えてよいだろうか。この問題を考えるために、同じ見出し語のサ変動詞としての用法を4)に例として取り上げよう。

- 4) 農夫がブドウを収穫する
農家が大豆を収穫する

「収穫する」のヲ格にたつ名詞句は、「ブドウ」や「大豆」などの作物を表す名詞句である。これらの名詞句は、1)、2)の「収穫」の連体修飾語として現れるものである。一方、以下の5)に示すように、3)の

「収穫」と共起する名詞句を組み合わせても、サ変動詞用法は成り立たない。

- 5) *議長が会議を収穫する
*今日の会議で知識を収穫する

従って、1)、2)の「収穫」と3)の「収穫」とでは、結びつく述語に同じものがあったとしても、修飾成分が異なり、その文法的ふるまいに違いがあることがわかる。よって「収穫」は次のように1)、2)と3)の二つに下位区分される。

1. 農作物をとり入れること。または取り入れた農作物など。 1) 2)
2. ある事をして得た有益な結果。 3)

4.3 「述語用法の有無」を用いた例

続いて、見出し語「田舎」の用例をあげる。

- 1) 田舎 に 生まれる／育つ
- 2) 田舎 が ある／ない
- 3) 田舎 へ 帰る／電話する

ここで、「とても」という程度副詞との共起関係を見ると、違いがはっきり見えてくる。

- 1) とても田舎 に 生まれる／育つ
- 2) *とても田舎 が ある／ない
- 3) *とても田舎 へ 帰る／電話する

つまり、「田舎」は、述語用法があるものとないないものに区分されることがわかる。ここでいう「述語用法」とは、次のような、程度副詞を伴って述語になる用法のことをいう。

- 4) 私が住んでいるところはとても田舎だ

結局、「田舎」は以下のように二つに下位区分される。

1. 都会から離れた所。 1) 4)
2. 故郷または出身地。 2) 3)

4.4 「結びつく述語と連体修飾語の違い」を用いた例

次に「点数」という見出し語について考察する。この「点数」という見出し語には、サ変動詞としての用法や述語用法がないため、前節であげた方法では下位区分することができない。このような場合に下位区分の判断基準となるのは、「結びつく述語の違い」と「連体修飾語の違い」である。

- 1) 応募作品の点数 が 多い／少ない
- 2) 絵画の点数 を 増やす／減らす
- 3) 数学の点数 が 高い／低い
- 4) 試験の点数 を 上げる／下げる

上の例では、1)、2)の「多い、少ない、増やす、減らす」と結びつく「点数」と、3)、4)の「高い、低い、上げる、下げる」と結びつく「点数」とでは、性格に違いがあると考えられる。また、それぞれの「点数」の連体修飾語にも、異なる性格の名詞句が現れている。1)、2)は修飾語に現れる名詞句が指示するものの個数を問題にするものであり、3)、4)は程度を問題にするものである。このような理由から「点数」については1)、2)と3)、4)の二つに下位区分することができる。

1. 試験やゲームなどの成績を表した数。 1) 2)
2. 商品・作品などの数。 3) 4)

4.5 「被修飾語の違い」を用いた例

「窮屈」という見出し語は、格助詞ガやヲを介して述語と結びつくという用法がない。また、他の名詞の連体修飾を受ける用法もない。あるのは次のようなa)連体用法、b)述語用法、c)格助詞ニを介して述語と結びつく用法である。

- a1) 窮屈なズボンをはきたくない
- a2) 窮屈な式典は肩がこる
- a3) 窮屈な制度は変えていくべきだ

- b1) ズボンが窮屈だ
- b2) 式典が窮屈だ
- b3) 制度が窮屈だ

- c1) ズボンが窮屈になる
- c2) 式典が窮屈になる
- c3) 制度が窮屈になる

そこで、a)連体用法の時に現れる「被修飾語の違い」に着目する。b)述語用法や、c)格助詞ニを介して述語と結びつく用法の時に違いとなって現れるのは、この「被修飾語」である。これを用いて下記のように下位区分する。

1. 狭くて、思うように身動きができないさま。 1)

[被修飾語] 一な服, 一な部屋, 一な姿勢
2. 堅苦しく、気詰まりに感じるさま。 2)

[被修飾語] 一な式典, 一な雰囲気, 一な感じ
3. ざりざりでゆとりがないさま。 3)

[被修飾語] 一な制度, 一な制限, 一な法,
一な金利体系, 一な暮らし, 一な生活

4.6 「合成語の違い」を用いた例

次の見出し語「神経」の例では、結びつく述語にも違いがあるが、「神経」を語基として成立する合成語の違いによっても下位区分できる。

1. 「しんけい」とよばれる身体器官。

例：神経 が麻痺する／を切る／を抜く
[合成語] 一細胞, 一中枢, 一痛, 視一, 自立一,
運動一, 中枢一, 脳一, 交感一
2. 「しんけい」の機能・働き。

例：神経 が弱る／を痛める／が過敏だ
[合成語] 一疾患, 一障害
3. 精神の働き、機能。

例：神経 が細い／細かい
図太い神経をしている
[合成語] 一安定剤, 一過敏, 一衰弱, 一戦,
無(む)一, 一質, 一性, 一症

5. 意味的な基準による下位区分

5.1 「意味カテゴリーの違い」を用いた例

ここで、意味カテゴリーというのは、「生物」「植物」「食べ物」「乗物」「衣服」「時間」「場所」などのような大きな意味分類を表す。見出し語「貝」は、「生物」「食べ物」「物」などの意味カテゴリーの違いで三つに下位区分できる。

1. 硬い殻をもち、水中に生息する「カイ」という生き物。

例：貝 が 棲む／プランクトンを食べる
貝 を 育てる
2. 「カイ」という生き物の食用部分。

例：貝 が おいしい／新鮮だ
貝 を 食べる／焼く
3. 「カイ」という生き物のからだを覆っている硬い殻。

例：貝 を 捨う／細工する／削る

5.2 「意味素性の組み合わせの違い」を用いた例

名詞と述語が格助詞を介して結びつく時、その述語によって名詞の持つ様々な側面の一つに焦点が当てられる。例えば、「手紙を破る」という時には、「手紙」の物質的な側面に焦点が当てられている。これに対して、「手紙を読む」では「手紙」の持つ内容的な側面に焦点が当てられている。IPA L名詞辞書における意味素性は、ある見出し語が持つ下位区分ごとに、結びつく述語によって焦点が当てられる側面とその範囲をとらえようとするものである[3]。

例えば、「みかん」という見出し語では、「ミカン」という植物そのものと、「ミカン」という植物の果実部分とを表す場合が考えられる。両者とも、具体物としての側面に焦点を当てる述語として「売る」「買う」のように、同じ動詞が共起するため、〈C O N〉という素性が振られる。さらに両者とも、植物としての側面に焦点が当てられる述語と共起するため、〈P L A〉という素性が振られる。しかし、「みかん」が植物を表す場合には、「食べる」「甘い」のような述語との共起は考えにくいので、〈E D I〉という意味素性は現れない。

両者の違いは、先にあげた意味カテゴリーの違い、「植物」「食べ物」の違いで下位区分することができるが、意味素性の組合せからも、〈EDI〉の有無によって次のように区分することができる。

1. 「みかん 1」「ミカン」と総称される植物。
 - 1.1 〈PLA:植物としての側面に焦点が当てられるもの〉
例：みかんを栽培する
みかんが枯れる
 - 1.2 〈CON:具体物としての側面に焦点が当てられるもの〉
例：みかんを買う
みかんを売る
2. 「みかん 2」「ミカン」と総称される植物の果実部分。
 - 2.1 〈EDI:可食物としての側面に焦点が当てられるもの〉
例：みかんを食べる
みかんが甘い
 - 2.2 〈PLA:植物としての側面に焦点が当てられるもの〉
例：みかんが生る
みかんが熟れる
 - 2.3 〈CON:具体物としての側面に焦点が当てられるもの〉
例：みかんを買う
みかんを盛る

5.3 「類義語や対語の組合せの違い」を用いた例

見出し語「悪質」の用例を挙げる。

- a1) 最近のスリの手口はますます悪質になってきている
- a2) 悪質な事件が多発している
- a3) 悪質な石炭を使わないようにする
- a4) 最近、悪質の酒が出回っているという

先ほどの「窮屈」という見出し語と同様に、格助詞ガやヲを介して述語と結びつく用法がなく、他の名

詞の連体修飾を受ける用法もない。よって、「被修飾語の違い」をみていくわけだが、この見出し語の場合、類義語の「粗悪」と対語の「良質」を用いることで違いがより明らかになる。

- b1) *そのスリの手口は粗悪だ/良質ではない
- b2) *その事件は粗悪だ/良質ではない
- b3) その石炭は粗悪だ/良質ではない
- b4) その酒は粗悪だ/良質ではない

つまり、類義語に「粗悪」を、対語に「良質」を持つものと持たないものとの、次のように下位区分される。

1. 行いがひどくたちの悪いさま。 1) 2)
 2. 品質が悪いさま。 3) 4)
- [類義語] 粗悪
[対語] 良質

5.4 「比喩的表現か否かの違い」を用いた例

「あの人は本の虫だ」の「虫」や「先日の講演会では講師が壇上で涙ぐむ一幕があった」の「一幕」のように、ある名詞がその指されたものの性質・特徴とよく似た性質・特徴を持つ別ものを表すことがある。これらは、比喩として使用されたと推測されるが、現代語ではかなり一般的に使われ、一つの用法として定着していると判断してもよいものである。【名詞辞書】では、このような名詞の表現を見出し語の持つ一つの用法として認め、下位区分としてたてた[4]。例えば、次の「虫」「卵」「種」「壁」「雨」「薬」「せのび」のようなものである。

- 1) 彼は本の虫だ
- 2) 医者のお卵と付き合う
- 3) 争いの種になる
- 4) 言葉の壁がある
- 5) 火の雨が降りそそぐ
- 6) 薬が効きすぎたようだ
- 7) 少女がせのびをして化粧をする

1)~5)では、連体修飾語の「NPの」部分が必須であり、これらの修飾語は、その見出し語が本来的な

意味を表していないことを明示している。これらは、修飾語の違いとしても下位区分できる。一方、6)~7)では、文脈の助けがないと本来的な意味との区別は難しい。このような場合には、形式的な基準を用いて下位区分することはできないが、意味的な違いを認めて別の下位区分をたてた。このように、比喩的な意味を一つの下位区分として認めた見出し語については、その比喩的な意味を表す下位区分と本来的な意味を表す下位区分とが大きく区別できるように区分した。

例えば、「雨」では、本来的な「雨」を表す下位区分と、5)で示したような「絶えまなくたくさん降りそそぐもの」という下位区分とに二分した。そのため、前者の「雨」には、多様な述語が共起することになるが、これらは、次のように意味素性を分けることによって細分される。

1. 「雨 1」広域に空から降ってくる水の粒。

それが降っている状態。

1.1 <CON: 具体物>

例：雨 がたまる／が冷たい／をよける

1.2 <LIQ: 液体>

例：雨 が漏る／がしみ込む／に濡れる

1.3 <PHE: 成立・消滅・変化する自然現象>

例：雨 が降る／にあう／になる

1.4 <NAT: 存在・移動する自然現象>

例：雨 が来る／が近づく／を予報する

2. 「雨 2」絶え間なくたくさん降り注ぐもの。

2.1 <CON: 具体物>

例：血／爆弾／花粉の 雨 が
降る／降り注ぐ／を浴びる

5.5 「総称か特化かの違い」を用いた例

「結びつく述語」に違いがあっても細かく下位区分をしない場合がある。例えば、見出し語「時計」の下位区分は次のように一つである。

1. 時刻を知るための機械。

例：時計 が 止まる／進む／遅れる

時計 を 見る／探す

実は、「時計」には形状によって次のように「結びつく述語」に違いがある。

- 1) (掛け) 時計 を かける／飾る
- 2) (置き) 時計 を 置く／飾る
- 3) (懐中) 時計 を 下げる
- 4) (腕) 時計 を つける／はめる

この違いで下位区分しなかった理由は、「時計」が「総称」を表すものであるからである。これを形状差によって下位区分すれば、辞書記述が冗長になる恐れがある。例えば、見出し語「コップ」を「コップが割れる」といえる「ガラス製」と、いえない「紙製」とで下位区分をするようなことにはあまり意味を感じない。自然言語処理において、形状が問題になることが予想されないわけではないが、IPALでは「総称」である名詞は、大きく用法を捉える方が良いと考え、形状差による下位区分は行っていない。同じ理由で、見出し語「薬」は、「飲み薬」「塗り薬」「貼り薬」で分けることをしていない。物としての側面に焦点が当てられる時には、「飲む」「塗る」「貼る」という違う述語と結びつくが、効果としての側面に焦点が当てられる時には、「効く」「作用する」「強い」「弱い」という同じ述語と結びつくことから、「薬」を形状で区分する必要がないことがわかる。

しかし、見出し語が「総称」を表す場合と、「特化」したものを表す場合とがある時には、その両者を下位区分した。以下に、「特化」で下位区分したものを三例示す。

見出し語「動物」：

1. 「ヒト」を含む生き物。

例：火星には動物がない

2. 「ヒト」をのぞく生き物。

例：サファリパークには色々な動物がいる

見出し語「ごはん」：

1. 食事。

例：今日は外でごはんを食べよう

2. 米を炊いて作る食物。

例：一日一膳は御飯を食べる

見出し語「酒」：

1. アルコールを成分として含む飲料、又、それが身体に及ぼす作用。

例：いろいろな酒が置いてある

2. 米を原料とする、日本に古くから伝わるアルコール飲料、又、それが身体に及ぼす作用。
例：酒を熱燗にする

上の例は三例とも、1が「総称」であり、2が「特化」である。

5.6 「類義語と対訳の違い」を用いた例

先に見出し語「応援」を下位区分した結果から示す。

- 1.1 他人の手助けをすること。〈ACT〉
例：選挙の応援を要請する
- 1.2 他人の手助けをするひと。〈HUM〉
例：たくさんの応援をよこす
- 2.1 声をかけたりしてチームや選手を元気づけること。〈ACT〉
例：チームの応援をする
- 2.2 声援を送っている人。〈HUM〉
例：たくさんの応援を呼ぶ

1と2は「結びつく述語」や「連体修飾語」などにそれほど差がなく、また、「意味素性の組合せ」も同じであるので、「相手を励まし、助ける行為」という下位区分を一つたてれば十分なかもしれない。しかし、以下のように「類義語」や「英訳」の違いに注意すれば、結局上のように二つに下位区分するのが適切である。

- | | |
|-----------|------------|
| a. 類義語が違う | 1. 支援, 手助け |
| | 2. 声援, 励まし |
| b. 英訳が違う | 1. help |
| | 2. cheer |

しかし、「対訳の違い」は、「下位区分の判断基準」の中でも適用度は低く、下位区分する絶対条件ではない。例えば、前節であげた「時計」は、形状によって以下のように英訳が分かれるものであった。

- 掛け時計：a (wall) clock
置時計：a (table) clock
懐中時計：a (pocket) watch
腕時計：a (wrist) watch

だが、先ほども述べた通り、「総称」であるものは一つの下位区分にするという方針にのっとり、このような場合の「対訳の違い」は考慮しないで、「時刻を知るための機械」という下位区分を一つたてるだけである。

6. おわりに

本稿で示したように、『名詞辞書』では11通りの「下位区分の判断基準」を設けたことによって、一貫した「下位区分」を行うことができた。見出し語の「下位区分」は機械的に行うことは難しい。IPAでは、辞書の仕様作成にあたった言語学・国語学の研究者が総合的な判断で「下位区分」をし、自ら辞書の記述と校閲を行うことで、辞書の記述内容の質の高さと均一性を保っているといえよう。

参考文献

- [1] 橋本三奈子, 桑畑和佳子, 村田賢一. 「IPA L名詞辞書の概要」. 『IPA Lシンポジウム '93 論文集』, pp. 1-12, 1993.
- [2] 桑畑和佳子, 橋本三奈子, 井口厚夫, 猪塚元, 村田賢一. 「IPA L名詞辞書におけるコロケーションの記述」. 情報処理振興事業協会『第13回技術発表会論文集』, pp. 73-76, 1994.
- [3] 橋本三奈子. 「名詞の意味素性と見出し語の下位区分」. 情報処理振興事業協会『第13回技術発表会論文集』, pp. 67-72, 1994.
- [4] 橋本三奈子, 桑畑和佳子, 青山文啓, 村田賢一. 「名詞の比喩的表現とその統語的特徴」. 『情報処理学会第49回全国大会論文集』, pp. 3-139, 1994.

謝辞

共同研究者である村田賢一氏、外池俊幸氏、青山文啓氏、山本清隆氏、井口厚夫氏、三枝令子氏、加藤久雄氏、加藤安彦氏、猪塚元氏、堤正典氏、小川裕花氏、高松正毅氏、中島尚樹氏、塩谷英一郎氏、渡辺恵子氏、鈴木高志氏、緒方典裕氏、渡辺和恵氏、木村朗子氏、斉藤初江氏、本多啓氏、窪田美穂子氏、宗意幸子氏、山下智弥氏、玉井陽子氏、長島みどり氏、他みなさんに深く感謝の意を表します。