

## 同形多品詞・同形多義語のための辞書設計

塚脇 幸代

s-tsuka[at]dream.ocn.ne.jp

## 概要

同じ形で複数の品詞が存在したり、複数の意味を持つ語をテキスト処理に適した形式で辞書に収めることを目的とし、名詞的接尾語を例にとりあげ、個々の名詞的接尾語がそれぞれに多様な読みと品詞と意味を持つことを観察しつつ、形(表記)をキーとした辞書構造を提案した。

## はじめに

日本語の品詞体系にあって扱いが難しい範疇の一つに接辞がある。『日本文法大辞典』によると、接辞とは「いつも他の語や語基に従属・融合・活用して、一語を構成する要素。それ単独では不安定で、体言や用言や副詞やそれらの語基についてはじめて一語をなし、語の内部で安定する。」(p.365)とある。第一義的にはそれが接続することにより品詞を変えたり、意味を添えたりするもので、形容詞+接尾辞「さ」「み」で名詞を形成したり(うれしい→うれしさ、たのしい→たのしみ)、丁寧さを付け加えたり(名前→お名前)といった例が想起されるが、言語処理においては、もう少し広い範囲を接辞として扱う習慣がある。「警察署」「消防署」の「署」、「保育園」「動物園」の「園」などがそれにあたる。これらは名詞的接尾辞(UniDicの品詞体系では「接尾辞-名詞的-一般」)などと呼ばれるが、名詞的接尾辞自体が語の構成要素のなかで中心的意味を担っており、本稿ではこれらを「接尾辞」ではなく「接尾語」と呼ぶ。また本稿で中心的に取り扱うのは接尾語としての用法と他の品詞用法を持つ一文字からなる「名詞的接尾語」である。

## 1. 名詞的接尾語と名詞・同義異音語・同形異義語

名詞的接尾語はその文字列が接尾語として用いられる以外に、単独で使用されることがある。この場合の品詞カテゴリは名詞となる。さらに名詞となった場合、読みが異なる場合がある。下に「米」の例を示す。まず(1-1)は名詞的接尾語であり、(1-2)の名詞の意味を持つ。

(1-1) 京都米(きょうとまい)【名詞的接尾語】

(1-2) 京都で米を作る。(きょうとでこめをつくる)【名詞】

また、他の意味・用法の同形の語彙が存在することがある。

(1-3) 日本政府は米の景気回復を期待した。(にはほんせいふはべい=あめりかのけいきかいふくをきたいした)【名詞・同形異義語】

(1-4) 太郎は百米走が得意だ。(たろうはひゃくめーとるそうがとくいだ)【助数詞-単位】

同形の見出しに対して複数の読みや品詞が与えられる場合は日本語において少なくない。ただ、与えられる読みの数や品詞の数、用法の違いは、それぞれの語によって異なるため、安易に一般化することができない。「米」が持つ読みの数と「園」や「市」が持つ読みの数は異なり、品詞の数や意味の数もそれぞれに異なる。個々の名詞的接尾語が持つ情報は、個別に記述されなければならない。

## 2. 名詞的接尾語の名詞用法にみる指示機能

名詞的接尾語が単独で使われると名詞的用法となる。

(2-1)の「園」(太字で表示)は「保育園」の省略として使われている。(2-2)の「園」も同じく「保育園」の省略とみられるが、保育園一般でなく、「ぼく」が滞在していた特定の「保育園」であると推測される。

(2-1) 幼稚園のお母さんたちにお聞きます。園の行事などで、私の着ている服を上から下までジロツと、あからさまに見てくる人がいます。(BCCWJ:OC10\_01035:140)【名詞-普通名詞-一般】

(2-2) ぼくは保育園の先生方に気どられぬようにそつと園を出た。(BCCWJ:PM52\_00059:43100)

さらに(2-3)において「市」とは単なる行政単位ではなく、「土地を買う」行為を行った特定の「市」を指している。

(2-3) あのキャンペーンは今でも有意義だったと確信している。ほかに、例えば、市役所前の一等地にあったマスコミ所有のビル売却問題。古いビルを更地にしないで、市が土地を買った。(BCCWJ:PM41\_01370:44800)

同様に、「署」とは国語辞典によると第一義に「消防署」「税務署」などの略、次に「警察署」の略とある。<sup>1</sup>(4)は警察署の略としての「署」である。ただし(5)においては同じく警察署のことではあるが、警察署一般ではなく、「新宿署」を指している。

(2-4) 「(中略)それより、首謀者の土佐平一郎という男をつかまなさい。あたしが監禁された家にたしかにいたんだから」「だから、その点について署でくわしく訊きたいと言っているんだ」(BCCWJ:PB39\_00436:29210)

(2-5) 鮫島は、渋谷の平出組事務所から直接新宿署に戻ることをせず、桃井だけを呼びだして事態を報告したのだった。署に戻れば、本庁から署長を通じて、圧力が降りている可能性があった。それを受ける時間を少しでも先延べするために、鮫島は署に戻らなかったのだ。(BCCWJ:OB5X\_00248:630,1010)

(2-4)は実際に出向く場所が不特定の警察署ではないにしろ、どこの「署」であるかは重要でなく、「警察署に行く」ことが焦点となっている。その意味ではこの「署」は「警察署」の省略であるという説明が成り立つ。しかし、(2-5)の例で「警察署」の略が「署」であると仮定すると、省略関係は同義関係であるという前提の下に、省略語「署」を「警察署」の具体例である「新宿署」に置き換えなければならない。(2-4)と同じく「警察署に行く」ことが焦点となっている(2-6)の例では、行き先が「網走警察署」であろうことが読み取れる。これでは個別の警察署名がすべて「署」の正式名称ということになり、省略関係で説明するには無理がある。

(2-6) 案の定、彼は内ポケットから黒い手帳を出して示した。「網走警察署の者です。尋ねたいことがありますから、署まで同行願います」(BCCWJ:LBt9\_00256:41380)

(2-5)や(2-6)の「署」が警察一般ではなく、「新宿署」あるいは「網走警察署」を意味するためには、「署」に指示の役割が必要となる。すなわち、「署」は既出の「新宿署」や「網走警察署」を指すと考える。既出のものを指し示すことができ、かつさまざまなクラスに属する対象を指すことができる。これはすなわち指示代名詞的な機能である。指示代名詞と異なるのは、「警察署」「税務署」「消防署」など、「署」で表すことのできる対象のみを指示することができる点である。あらかじめ対象が制約された指示代名詞と行うことができる。「園」や「市」にも同様の機能が備わっていると考えられる。「園」であれば「動物園」「植物園」「保育園」「幼稚園」などの一般名称と「旭川動物園」「ひまわり幼稚園」などの個別名称を指示することができる。

(2-7) うちの幼稚園はお休みしたり、園の都合で午前保育だったりした場合でも、毎月決まった金額を払います。(BCCWJ:OC10\_04592:220)

一方「市」の場合は、「大阪市」「札幌市」のような個別名称を指示することができるが、「政令指定市」などの一般名称を指示することはできない。この機能においてもそれぞれの語に適した機能をそれぞれに記述する必要がある。

### 3. 形と音と品詞と意味の関係

<sup>1</sup>株式会社小学館発行『デジタル大辞泉』©Shogakukan による。

すでにみたように、名詞的接尾語の中には多様な音(読み)と多様な品詞を持つものがある。ある音形とある品詞、またはある意味との結びつきもまた、それぞれの語によって異なっている。「米」という形に対し、「まい」という音が与えられたとき、それは名詞以外の品詞であり(京都米/古米)、「こめ」という音が与えられた場合には、接尾語ではない可能性が高くなり、「べい」という音であれば、固有名(国名)や穀物名を想起させる。一方、穀物名であっても形が「コメ」であれば、「コメント」の略と解釈される可能性も出てくる。どの音とどの形が結びつき、どの品詞と結びついていくのかという問題もまた、それぞれの語に対し、個別に解決されるべきである。形から音(読み)、さらに品詞や意味への結びつき方はたとえば以下のように示される。<sup>2</sup>

#### 形→音

米→こめ/まい/べい/めーとる/よね

園→えん/その

市→いち/し

#### 形→音→品詞

米→こめ→名詞 [米を作る]

米→めーとる→助数詞 [10米]

米→べい→名詞/不定:語の一部 [米の景気回復/米飯]

米→まい→名詞的接尾語/不定:語の一部 [京都米/米原]

米→よね→名詞/不定:語の一部 [お米さん/米坂線]

園→その→名詞 [エデンの園]

園→えん→名詞的接尾語/指示代名詞 [保育園/園]

市→いち→名詞/名詞的接尾語 [蚕の市/見本市]

市→し→名詞/指示代名詞 [2006年に市になった/市が学校を建て替える]

#### 形→意味

米→穀物/国名/単位 (複数の意味に対応)

また、異表記関係や同義関係、省略関係などは形から形への対応と言える。

異表記や同義関係まで展開されれば、上位下位関係などの、他の関係概念ネットワークへ接続することができる。

#### 形→形

米→コメ (異表記)

米→アメリカ (同義関係)

コメ→コメント (省略関係)

コメント→注釈 (同義関係)

コメ→米 (異表記)

#### 4. 形をキーとする辞書構造

音(読み)を左端(=キー)に持つ辞書構造は、既存の辞書の基本的な構造と一致する。音声を入力とする場合や、かな漢字変換処理を行う場合にも、音(読み)は重要な手掛かりとなる。しかし、テキスト処理においては、テキストにおいて使われている形(=表記)が入力となるため、形をキーとする構造が適している。読みをキーとする辞書は、読めなければ引けないという弱点がある。キーとなる形には、漢字に限らず、実際のテキストに使われ

<sup>2</sup> どの単位を一語と扱うかによって、展開される要素の種類は変化する。

る表記であれば当然ひらがなやカタカナも採用できる。

## 5. 曖昧性を前提とした辞書構築

形をキーとした辞書構造を採用しても、品詞を決定する際に負荷が軽減されたり、多義や曖昧性が解消されたりするわけではない。むしろ負荷を増大させる恐れさえある。それでも、人の目が見分けられる範囲の見出しを、辞書は保持してはならない。処理すべき対象の姿をできるだけ忠実に収めておくことが辞書の役割である。どの表記がどの読みで使われ、どのような意味で使われているのか、どの分野でどの表記が主に使われ、どの意味が優先されるのかといった情報は、十分に整備されたコーパスデータから統計処理等の力を借りて、いずれ明らかにされていくと考える。

## おわりに

複数の読みや品詞、同形多義を持つ名詞的接尾語を例に用いて、形をキーとした辞書構造を提案した。言語の曖昧性は言語処理を行う際に多大な苦勞を与えてきた。曖昧性を解消するために、様々な工夫や努力がなされてきたが、その過程で処理を優先するために、言語事実の記述を犠牲にしてきた部分があったかもしれない。逆に辞書や言語資源を構築する側にあつては、言語処理のあり方を見極められないままデータを構築してきたきらいもある。言語処理と辞書は、互いに依存関係にある。信頼性の高い辞書を構築すれば、言語処理の精度も上がっていくと期待する。

## 謝辞

本稿に記載した例文のうち文末に BCCWJ と表記したものは、大学共同利用機関法人人間文化研究機構国立国語研究所コーパス開発センターの『現代日本語書き言葉均衡コーパス』から検索ツール「中納言」を使用して抽出した。ここに記して感謝の意を表す。

## 参考文献

- [1] 松村明編. (1971). 『日本文法大辞典』. 明治書院.
- [2] 大学共同利用機関法人人間文化研究機構国立国語研究所コーパス開発センター. (2011). 「『現代日本語書き言葉均衡コーパス』利用の手引」第 1.0 版.
- [3] 伝康晴, 小木曾智信, 小椋秀樹, 山田篤, 峯松信明, 内元清貴, 小磯花絵. (2007). 「コーパス日本語学のための言語資源: 形態素解析用電子化辞書の開発とその応用」. 『日本語科学』22 号 pp.101-122.