

辞書編集における見出し語の選定に関する一考察

齊 藤 康 彦

北海道情報大学

A Study of Word List Selection in Dictionary Making

Yasuhiko SAITO

Hokkaido Information University

平成29年 3 月

北海道情報大学紀要 第28巻 第 2 号別刷

〈研究ノート〉

辞書編集における見出し語の選定に関する一考察

斉藤康彦*

A Study of Word List Selection in Dictionary Making

Yasuhiko SAITO*

要旨

収録された見出し語の語釈を解析することによって、辞書編集における見出し語の適切な選定を支援する手法を提案する。本手法では、語釈の理解を目的としたときの基本語、すなわち、必要最小限の基礎的な概念を抽出し、それらが見出し語として当該辞書に収録されているかどうかを確認する。さらに、当該辞書における見出し語の選定にどのような特徴があるかを把握する。このとき、基本語であるかどうかの基準が選定者に依存することが問題となるが、本手法では、その語の語義木中での出現回数に基づいて、系統的に基本語を抽出する。語義木とは、語と、その語の語釈中で用いられている語の間に定義される親子関係に基づく木構造である。本稿では、幼児向けの国語辞典に本手法を適用する。

Abstract

A method of analyzing definitions of words in a dictionary is proposed. The purpose of this method is to examine word list selection in dictionary making. Using this method, we can extract fundamental words representing minimum basic concepts from a dictionary. These words are important for interpreting definitions of words in the dictionary. It is possible to check whether these fundamental words are selected as entries of the dictionary. Also, we can grasp characteristics regarding word list selection of the dictionary. In many cases, which words are chosen as fundamental words depends on who chooses them. The proposed method, however, allows us to extract fundamental words systematically according to how many times each word is used in whole of *semantic trees*, each of which is constructed from parent-child relationships among words used in the definition of each word in the dictionary. In this report, the method is applied to a Japanese dictionary for young children.

キーワード

辞書学 (lexicography) 辞書批評 (dictionary criticism) 語義 (semantics) 基本語 (fundamental words) 基礎語 (basic words)

* 北海道情報大学経営情報学部システム情報学科准教授, Associate Professor, Department of Business and Information Systems, HIU

1. はじめに

言語は、社会生活を営む上で必要不可欠なものである。辞書には、それぞれの時代に、それぞれの社会で使用されてきた／されている言語が記録されている。したがって、その時代、その社会の文化を、辞書から読み解くことができる。現代社会において、コンピュータとネットワークを基盤とした文化は、辞書の形態と内容に大きな変化をもたらしている。それは辞書の電子化である。

占来、辞書は、紙の上に文字で記されてきたが、電子辞書は、コンピュータ内部や電子媒体に電子データとして保存される。こうした形態の電子辞書には、紙の辞書と同じように人間が使うためのものと、情報検索や機械翻訳などをするためにコンピュータシステムが使う（アクセスする）ためのものがある。ここでは、人間が使うための辞書を、コンピュータを利用して編集するという観点から、辞書の電子化について考えてみたい。

人間が使うための辞書には、以下の種類がありえる。

- (a) 紙に印刷された辞書
- (b) 紙に印刷された辞書をほぼそのまま電子化した辞書
- (c) 紙に印刷された辞書を大幅に拡張して電子化した辞書
- (d) はじめから電子化されるものとして編集された辞書

(b)と(c)は、まず紙の辞書が編集され、それが電子化される場合であるが、(c)では、紙の辞書では避けられなかった紙面スペースの制約がなくなる。(d)の事例としては、Wiktionaryが挙げられるが、これも紙面スペースの制約がない。したがって、論理的には、いくらでも多く見出し語を収録することができて、いくらでも長く語釈を書くことができるようになる。紙の辞書の場合でも、紙面スペースの

制約を考慮しないで編集することは可能である。しかし、たとえば、全14巻からなる最大規模の紙の国語辞典である小学館の『日本国語大辞典』は、日常的な使用には適さない。

紙面スペースの制約がなくなることは、電子化の恩恵である。しかし、一方では、紙の辞書の利点も指摘されている。岩波書店の国語辞典『広辞苑』などの編集に携わった増井によれば、紙の辞書には、電子辞書に対して以下の利点がある[増井(2013)]。

一つは、読みやすさです。「文字・文字の大きさやデザインのことでなく、紙の辞書の方が項目全体の見通しがきくということです。改良はされて行くでしょうが、電子辞書の画面は狭いのです。もちろん、その時点で読んでいる箇所はよく見えるのですが、全体の記述の中で、どんな位置をしめている記述かが掴みにくく感じられます。そうした言わば「一覧性」とでもいう点で、電子辞書は劣っています。

これこそ紙の辞書ならではと思うのは、やはり目が届く範囲の広さと関係することですが、いま探しているのではない項目に目が行くということです。そんな直接必要でない周りの項目が見えてどうなるのだ、と言われても困るのですが、これはなかなかのもので、そこにはことばや辞書と付き合うときの喜びが隠れているのです。

さらに、紙面スペースの制約から語釈の長さを短くせざるを得ない点に関しては、「長く詳しく説明することは辞書の手柄にはならない、一言で言ったら何なんだ」という先輩からの忠告を、辞書編集の中核となる心得であると述べている[前掲書]。

また、三省堂の国語辞典『三省堂国語辞典』

の編集に携わった飯間は、紙の辞書に収録される見出し語の数が制限されることによって、収録される見出し語に辞書ごとの違いが生じるが、辞書の編集者としては、他の辞書に収録されていない語が、自分が編集する辞書に収録されていることに喜びを感じると述べている[飯間(2013)]。このことは、辞書の利用者にとっては、複数の辞書を比較して、各々の個性の違いを楽しむことに繋がると思われる。

そこで、本研究では、紙面スペースの制約があることを肯定的に捉えて、可能な限り多くの見出し語を収録するのではなく、収録する見出し語を選定することを前提とする。その際、編集の対象とする辞書は、紙の辞書か電子辞書かを問わない。

本稿では、辞書編集において見出し語を適切に選定することを支援するために、辞書に収録された見出し語をコンピュータを利用して解析する手法を提案する。本手法では、辞書に収録すべき基本語の候補を抽出する。基本語とは、さまざまな語の意味を理解する上での基礎となる重要な語である。本手法によって、基本語であるにも関わらず、見出し語として収録されていない語が検出できるので、必要であれば、その語を見出し語として収録することが可能になる。また、本手法によって、辞書に収録する語の選定に関する特徴を把握できる。その結果として、収録する語に偏りが認められ、それが意図した偏りでない(不適切である)場合は、収録する語の見直しが可能になる。

2. 辞書編集における見出し語の選定

一口に辞書といっても、さまざまな種類がある。国語辞典、漢和辞典、英和辞典、和英辞典、英英辞典、専門用語辞典、百科事典などがあるが、本稿では、国語辞典を対象とする。一般的な国語辞典には、日本語の語について、日本語を用いて、見出し、漢字表記、

品詞、語釈、用例などが記述されている。

三省堂の国語辞典『大辞林』の編集に携わった倉島によると、辞書編集の過程は次のようなものである[倉島(2008)]。まず、基本構想、基本調査、企画立案などの、準備・計画作業を行い、それを踏まえて、順次、「編集方針」、「編集要領」、「執筆要領」を作成する。その上で、執筆作業が開始される。執筆においては、まず、項目選定、すなわち、見出し語として収録する語を選ぶ作業が行われる。項目選定に先立って、語彙カードと呼ばれる情報を事前に用意しておく。語彙カードとは、収録する見出し語の候補について、その用例を集めて記載したものである(紙のカードに限らず、コンピュータに入力されたデータも含む)。この語彙カードに基づいて収録する語を選ぶのであるが、ここで注意すべきことは、この段階では、まだ語釈が書かれていないので、用例に基づいて選ぶという点である。項目選定が終わると、各項目の内容を執筆していく。その中心となるのが語釈である。語釈は、辞書の評価を大きく左右するものである。しかし、辞書の利用者からすれば、調べたい語が収録されていなければどうにもならないので、収録する見出し語を適切に選ぶことは、辞書編集の過程において極めて重要である。

岩波書店の国語辞典『岩波国語辞典』の編集に携わった水谷によると、見出し語の選定では、バランスが大切である[水谷(2012)]。バランスがよいというのは、基本語を見出し語として収録している状態である。基本語とは、現代人が生活の中で普通に読み書きするのに必要な語である。たとえば、ある語の意味を調べるために辞書を引いたときに、その語釈で使われている語が理解できなければ、当該語の語釈が理解できないことになる。したがって、その場合に、語釈で使われている語の多くは基本語であり、それらの基本語を見出し語として網羅することで、バランスのよさが生まれる。逆に、見出し語として収録され

ていない基本語が多いければ、見出し語を偏って選定している可能性がある。

ところで、ある語が基本語であるかどうかの判定は、それほど容易ではない。以下のように、基本語の選定基準が一通りでなく、しかも、主観的な要素を含むからである。

基準1

特定の調査対象における語の出現回数に基づいて選定するもので、最も客観的な基準である。

基準2

日常の言語生活を営む上で必要最小限の基礎的な概念を言語の専門家などが選定するもので、選定者の主観に依存する基準である。

基準3

何らかの目的を想定し、その目的を達成するための初期段階に必要な語を選定するものである。したがって、選定される語は、設定した目的によって変わってくる。代表的な目的の例として、学校での母国語の教育や、外国人のための日本語教育などがある。

これらの基準は、独立して適用されるものではない。水谷は、基本語が備えるべき性質として、以下を挙げている[水谷(1983)]。

- 1) その語を封じたら、他の語での代用が利かず従って文章が綴れないか、他の語での代用がしにくく強いて言い換えるとかえって不便であるかである。
- 2) それらの語の組合せで、他の複雑な概念や新たな命名が必要な概念などが作りやすい。また現に、そうして作られた複合語が沢山在る。
- 3) それら以外の語の説明が必要な時にも、結局はそれらの語の範囲で行える。
- 4) そういう語の多くは、昔から使われ

て来たり、将来もまた使い続けられるであろう。

- 5) その語は多方面の話題を通じて割合よく使われる。

ここで、1)と2)と3)は、主に基準2によるものであるが、3)には、辞書の語釈を目的とするような観点も認められるので、基準3によっている。4)と5)は、出現回数の通時的・共時的な広がりやを考慮しているといえるので、基準1によっている。

本研究では、辞書の内容を解析することによって、語釈の理解を目的としたときの基本語を抽出し、それらが見出し語として当該辞書に収録されているかどうか、さらに、当該辞書における見出し語の選定にどのような特徴があるかを考察する。本研究における基本語は、語釈の理解を目的とする点で、基準3によるものであるが、語の出現回数を解析の対象としている点で、基準1にもよっている。

3. 語義木を用いた辞書の解析

3-1 対象とする辞書

本稿では、『三省堂 こどもことば絵じてん』[金田一(2009)]を対象に基本語の抽出を試みた。さらに、その結果に基づいて、本辞典の特徴を考察した。

本辞典は、6歳ぐらいまでの幼児に対して、大人が言葉の意味を説明するときを使うことを想定した、幼児向けの国語辞典である。一般の国語辞典と同様に、見出し語が50音順に配列されている。各見出し語について、その意味を表すイラスト、簡単な語釈、用例が掲載されている。本研究では、イラストや用例は無視して、語釈のみ扱う。語釈は、平仮名のみによる記述で、文節ごとに空白で区切られている。本辞典には約2900語が収録されているが、その中には、語釈が記載されていないものもある。今回の解析では、そのような

語を除いた 2411 語を対象とした。

本辞典を選んだ理由は、収録語が少なく、語釈が簡略であるため、解析に必要なデータを準備するためのコストが削減できると考えたからである。一般の国語辞典を対象とする場合には、データの準備に膨大なコストがかかる。本研究は、新しい解析手法の模索段階にあるので、まず手始めに、小規模なデータによる実験的な解析を試みることにした。

3-2 基本語と辞書の語釈

本研究では、「見出し語の語釈を理解する」という目的を想定した上で必要となる語が基本語であると考えられる。たとえば、本辞典における見出し語「りよかん」の語釈は、「おかねを はらって とまる ところ」となっているが、この語釈を理解するためには、「おかね」、「はらう」、「とまる」、「ところ」という 4 語の意味が分かっている必要がある。したがって、この 4 語は基本語の候補となる。しかし、直ちに基本語と認められる訳ではない。この語釈を読んだ者は、この 4 語（のいずれか）の意味が分からないこともあるが、その場合には、たとえば、「はらう」の意味が分からなければ、「はらう」を見出し語として辞書を引くことになる。このようにして、不明な語の意味を調べるために辞書を引くことを繰り返していき、もうそれ以上、辞書を引く必要がなくなった時点での語が、本辞典の語釈を理解する上で必要最小限の基礎的な概念に相当する語であると考えられる。すなわち、本研究における基本語とは、語釈の中で使われている語のすべてではなく、その中でも特に基礎的な語である。

ここで問題となるのは、「必要最小限の基礎的な概念」としての基本語の基準は、選定者の主観に依存する点である。本研究では、客観的なデータを解析することで、主観に依存することなく、系統的に基本語を抽出することを目的とする。そこで、その語の出現回数

に基づいて、必要最小限の基礎的な概念であるかどうかを判定することを試みた。これまでも、語の出現回数に基づいて、基本語を選定することが行われてきた[国立国語研究所(1982)][国立国語研究所(2009)]。本研究では、対象とする辞書の語釈中での回数ではなく、次節に詳細を示す語義木の中での回数に着目している点が、従来の選定の考え方と異なる。語義木とは、語と、その語の語釈中で用いられている語の間に定義される親子関係に基づく木構造である。語義木と同じような考え方によって、辞書の語釈における基本語を選定することが、水谷によって試みられている[水谷(2012)]。ただ、その際に、語の出現回数には注目していない。

3-3 語義木

3-3-1 定義

[辞書]

辞書は、見出し語の集合である。日本語、英語など、単一の言語を対象とする。

[語]

語 w は、語の識別子 $label$ と語義集合 S の順序対 $(label, S)$ である。 $label$ は、語 w を一意に識別する文字列である。語の中で辞書の要素であるものが見出し語である。

[語義集合]

語 w の語義集合は、語 w の語義要素の集合である。 \emptyset のこともある。

[語義要素]

語 w の語義要素は、語 w の語釈、すなわち、当該辞書が対象とする言語で語 w の意味を記述したものから、所定の規則にしたがって抽出した語である。

[語義木]

語 w の語義木 t は、語 w と語 w の副語義木

```

procedure get_child(X):
begin
  T ← ∅;
  S ← X の語義集合;
  for each e in S begin
    t ← (e, get_child(e));
    T の要素として t を追加する
  end;
  return T
end

```

図1 副語義木集合を得る手続き

集合 T の順序対(w, T)である。

[副語義木集合]

副語義木集合は、語義木の集合である。語 w の副語義木集合 T は、図1の手続き `get_child` を、引数に w を指定して実行することによって得られる。∅ のこともある。

[語義木における語の親子関係]

語 w と、w の語義要素とは、親子関係にある。すなわち、語義木(w, T)において、語 w が親で、副語義木集合 T の要素である語義木(w_e, T_e)の語 w_e が子である。このように、語義木は、語をノードする親子関係によって構成された木構造である。

[語義木の階層]

語義木における親子関係で、親の階層を n とするとき、子の階層を n+1 とする。木のルート階層を 0 とする。

3-3-2 語義木の生成における制約条件

語義木において、ある語が、その語の子孫に再び出現した場合、階層が限りなく深くなる無限循環に陥る。そこで、この場合には、当該語が再び出現したノードについて、副語義木集合を ∅ とする。これによって、無限大の語義木の生成を回避することができる。

語義木では、副語義木集合が空である語のノードが木のリーフになる。しかし、リーフの階層が深すぎる場合には、便宜的に、指定した階層にあるノードの語について、副語義木集合を ∅ とすることができる。これによって、指定した階層よりも深い階層を省略した（すなわち、指定した階層までの）語義木を生成することができる。

3-3-3 語義木の例

図2は、見出し語「りよかん」の語義木 t を定義したものである。階層を 5 までに指定する。図3は、語義木 t における語の親子関係を視覚的に表現したものである。図3において、木のリーフにつけられた記号<R>と<+>は、以下を意味する。

<R> このノードの語は、その語の祖先に既に出現しており、ここに再び出現した。

<+> このノードの語には子が存在するが、指定された階層を超えてしまうので、このノードより先に木を生成しない。

4. 語義データベース

4-1 語義データベースに登録する情報

語義木は、語義データベースに基づいて生成される。本データベースには、見出し語ごとに、見出し、表記、語義要素を登録する。

見出しは、見出し語の読みを平仮名のみで記述したものである。

表記は、漢字、平仮名、片仮名などを組み合わせて、見出し語の表記を記述したものである。同音異義語を識別する役割もある。たとえば、本辞典では、「まどを あける」の「開ける」と、「よるが あける」の「明ける」を、異なる見出し語として扱っているが、見出し語の「あける」だけでは識別できないので、漢字表記の「開ける」と「明ける」で識別する。また、「せが たかい」の「高い」と、「た

t=(“りよかん”, {e1, e2, e3, e4}), {t1, t2, t3, t4}	t3211 (c3211, ø)
t1=(e1, ø)	t3212=(e3212, ø)
t2=(e2, {t21, t22, t23})	e3211=(“わたしたち”, ø)
t3=(e3, {t31, t32, t33, t34, t35, t36})	e3212=(“こと”, ø)
t4 (e4, ø)	t3221=(e3221, {t32211})
e1 (“おかね”, ø)	t3222 (c3222, {t32221, t32222, t32223})
e2 (“はらう”, {e21, e22, e23})	t3223=(e3223, ø)
e3=(“とまる[泊まる]”, {e31, e32, e33, e34, e35, e36})	t3224 (c3224, ø)
e4=(“ところ”, ø)	c3221 (“ばしよ”, {c32211})
t21=(e21, ø)	c3222 (“きめる”, {c32221, c32222, c32223})
t22 (e22, {t221, t222, t223, t224})	e3223=(“くらす”, ø)
t23=(e23, ø)	e3224=(“こと”, ø)
e21=(“おかね”, ø)	t2221=(e2221, ø)
e22=(“だす”, {e221, e222, e223, e224})	t22212 (c22212, ø)
e23 (“こと”, ø)	t22213=(e22213, ø)
t31=(e31, {t311, t312, t313})	t22214=(e22214, ø)
t32=(e32, {t321, t322, t323})	c22211 (“うら[裏]”, S22211)
t33 (c33, {t331, t332})	c22212 (“ほう”, ø)
t34=(e34, ø)	e22213=(“いえ”, S22213)
t35=(e35, ø)	e22214=(“そと”, ø)
t36 (c36, ø)	t22321=(e22321, ø)
e31 (“じぶん”, {e311, e312, e313})	t22322 (c22322, ø)
e32=(“いえ”, {e321, e322, e323})	t22323=(e22323, ø)
e33=(“ない”, {e331, e332})	t22324=(e22324, ø)
e34=(“ところ”, ø)	t22325 (c22325, ø)
e35 (“ねる”, ø)	e22321=(“ある”, ø)
e36 (“こと”, ø)	e22322=(“ばしよ”, S22322)
t221=(e221, {t2211})	e22323=(“ちがう”, S22323)
t222=(e222, {t2221})	e22324=(“うつす[移す]”, ø)
t223 (c223, {t2231, t2232, t2233, t2234, t2235, t2236})	c22325 (“こと”, ø)
t224=(e224, ø)	t22331=(e22331, ø)
e221=(“なか[中]”, {e2211})	t22332=(e22332, ø)
e222=(“そと”, {e2221})	t22333 (c22333, ø)
e223 (“うつす[移す]”, {e2231, e2232, e2233, e2234, e2235, e2236})	e22331 (“おなじ”, S22331)
e224=(“こと”, ø)	e22332=(“こと”, ø)
t311 (c311, ø)	e22333=(“ただしい”, S22333)
t312=(e312, ø)	t22341 (c22341, ø)
t313=(e313, ø)	c22341 (“ところ”, ø)
e311=(“わたし”, ø)	t22351=(e22351, ø)
e312 (“こと”, ø)	t22352=(e22352, ø)
e313 (“ぼく”, ø)	t22353 (c22353, ø)
t321=(e321, {t3211, t3212})	c22351 (“のせる”, S22351)
t322=(e322, {t3221, t3222, t3223, t3224})	e22352=(“こと”, ø)
t323 (c323, ø)	e22353=(“すえる”, ø)
e321 (“ひと”, {c3211, c3212})	t32211=(e32211, ø)
e322=(“すむ[住む_棲む]”, {e3221, e3222, e3223, e3224})	e32211=(“ところ”, ø)
e323=(“たてもの”, ø)	t32221 (c32221, ø)
t331=(e331, ø)	t32222=(e32222, ø)
t332 (c332, ø)	t32223=(e32223, ø)
e331 (“からっぽ”, ø)	e32221=(“はっきり”, S32221)
e332=(“ようす”, ø)	e32222=(“する”, S32222)
t2211=(e2211, ø)	c32223 (“こと”, ø)
e2211=(“うちがわ”, ø)	
t2221 (c2221, {t22211, t22212, t22213, t22214})	
c2221 (“おもて”, {c22211, c22212, c22213, c22214})	
t2231=(e2231, ø)	
t2232=(e2232, {t22321, t22322, t22323, t22324, t22325})	
t2233 (c2233, {t22331, t22332, t22333})	
t2234=(e2234, {t22341})	
t2235 (c2235, {t22351, t22352, t22353})	
t2236=(e2236, ø)	
e2231=(“もの”, ø)	
e2232=(“うごかす”, {e22321, e22322, e22323, e22324, e22325})	
e2233 (“ちがう”, {e22331, e22332, e22333})	
e2234=(“ばしよ”, {e22341})	
e2235=(“おく[置く]”, {e22351, e22352, e22353})	
e2236=(“こと”, ø)	

図2 「りよかん」の階層5までの語義木

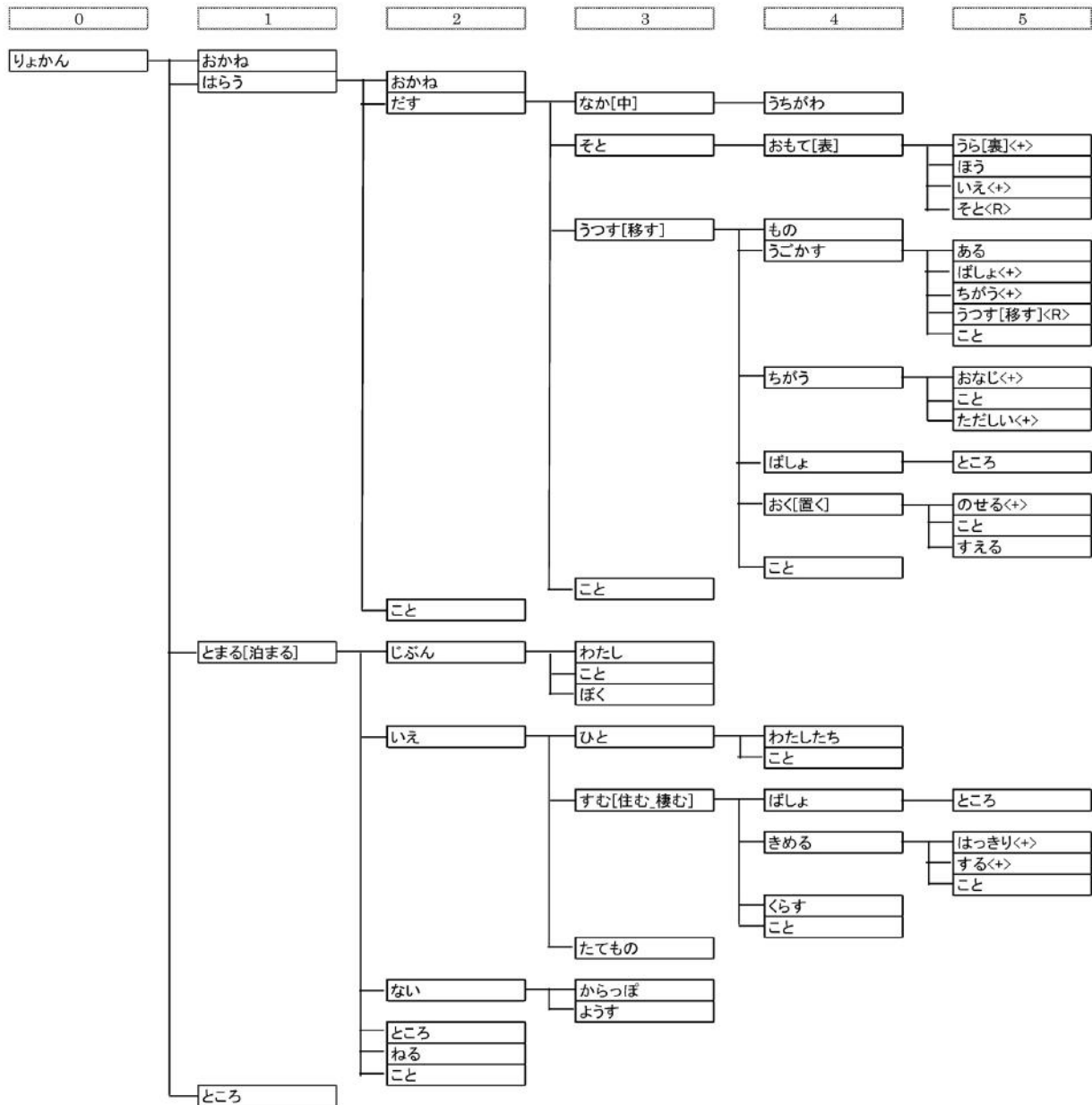


図3 「りよかん」の階層5までの語義木の視覚的表現

かい おもちゃを かって もらう」の「高い」を、異なる見出し語として扱っているが、この場合は漢字表記も同じなので、「高い1」「高い2」のように数字を付加して識別する。本研究では、同音異義語については、「あける[開ける]」のように、見出しの後の[]内に表記を付加したものを見出し語としている。

語義要素は、各見出し語の語釈に対して形態素解析を行い、自立語を抽出したものである。本辞典では、語釈がすべて平仮名で記述されているため、既存の形態素解析ソフトウェアによる形態素解析の自動化はしなかった。

人手で形態素解析を行った上で、必要に応じて、形態素解析ソフトウェアの MeCab[MeCab公式サイト(2016)]を使って、解析結果をチェックした。

4-2 語義要素の抽出

見出し語によって、語義要素の抽出方法に違いがあると、適切な語義木が生成できない。本稿では、以下の要領で語義要素を抽出した。

- 同一の見出し語の語釈中に2回以上出現する語は、1回(1語)だけ登録する。

- 助詞や助動詞などの付属語は登録しない。
- 以下のような、用言に付加される形で使われる語は登録しない。
 ～てあげる、～てある、～ていく、
 ～ている、～うち、～ておく、～がる、
 ～てくる、～てくれる、～てしまう、
 ～そう、～たび、～ため、～とおり、
 ～にくい、～まま、～てみせる、～てみる、
 ～てもらう、～やすい、～よう
- 動詞や形容詞など、活用がある語(用言)は、終止形で登録する。
- 以下のような、形容詞+「な」の形の連体詞は、形容詞の形で登録する。
 おおきな→おおきい
- 以下のような、形容詞+「さ」の形の名詞は、形容詞の形で登録する
 おもさ→おもい
- 以下のような、動詞の名詞形は、そのまま名詞として登録する。
 あそび、あつまり、いたみ、うごき、
 おどり、おわり、こおり、ながれ、はじめ、
 まちがい、もりあがり
- 以下のような、数に付加された形で使われる単位に相当する名詞は、数と切り離して登録する。その際、「～えん[～円]」のように、数の部分を「～」で示す。
 円, 回, 階, 個, 冊, 台, 度, 人, 杯, 本, 枚, 羽
 なお、特別なケースとして、「何度」「何枚」のように、数の部分が「何」になったものも同様に扱う。
- 語頭に敬語の「お」「ご」などがある場合は、それを付けたまま登録する。

- 複合語は、原則として、複合語のまま登録する。ただし、以下のような、助詞によって結合された複合語は、複合語を分割して登録する。
 おんなのこ→おんな+こ、
 はらをたてる→はら+たてる、
 きになる→き+なる、
 てにいれる→て+いれる、
 みにつける→み+つける、
 やくにたつ→やく+たつ、
 なんとなく→なに+ない

5. 全語義木中での出現回数に基づく基本語の抽出

全語積中での出現回数の多い語義要素は、基本語であるとは限らない。本研究では、全語義木中での出現回数の多い語義要素の方が、より基本的であると仮定する。たとえば、図2の「りよかん」の語義木において、出現回数が3回以上の語義要素は、「こと (11回)」「ところ (4回)」「ばしょ (3回)」である。これらの語は、直観的には、さまざまな語の意味を理解する上で共通の前提となるものと思われるが、「りよかん」以外の見出し語においても出現回数が多ければ、この辞書に収録されている見出し語の語積を理解する上で基本となる語であるとみなすことができる。そこで、本研究では、全語積中での出現回数ではなく、全語義木中での出現回数に基づいて、基本語を抽出することを試みる。

語 w の全語義木中での出現回数 N は、図4の手続き count を、引数に w を指定して実行することによって得られる。

このようにして、全語義要素についての全語義木中での出現回数を求めた上で、以下に示す手順で基本語を抽出した。

Step 1

図5は、木の階層が1～5の語義木について、階層が n から $n+1$ に変わることで、各階層に

```

procedure count(X):
begin
  H ← 辞書;
  N ← 0;
  for each h in H begin
    S ← h の語義集合;
    N ← N + count_sub(X, S)
  end;
  return N
end

procedure count_sub(X, Y):
begin
  N ← 0;
  for each e in Y begin
    if e=X then N ← N + 1;
    if e の副語義木集合 ≠ ∅ then begin
      S ← e の語義集合;
      N ← N + count_sub(X, S)
    end
  end;
  return N
end

```

図4 全語義木中での出現回数を得る手続き

おける全語義木中での出現回数の多い上位 30 語の順位がどのように入れ替わるかを示したものである。

階層 1 の全語義木中での出現回数は、全語積中での出現回数と一致する。階層 1 から階層 2 に変わる場合、順位変動の幅が大きく、階層 1 での上位 30 語が、階層 2 では、65 位までの範囲に幅広く散らばっている。階層 2 から階層 3 に変わると、順位変動の幅がやや小さくなり、階層 2 での上位 30 語が、階層 3 では、58 位までの範囲に散らばる。階層 3 から階層 4 に変わると、順位変動の幅がさらに小さくなり、階層 3 での上位 30 語が、階層 4 では、46 位までの範囲に散らばる。階層 4 から階層 5 に変わると、階層 4 での上位 30 語が、階層 5 でも、37 位までの範囲に収まるように

なる。順位があまり変わらない語も多い。

このように、階層が深くなるにつれて、順位が安定していく傾向が認められる。そこで、どれぐらいの階層まで調べれば、ほぼ安定した順位が得られるかが問題になる。必要以上に深すぎる階層まで調べることは、基本語の抽出という本来の目的から外れることになるので、語義木の階層を、順位が安定する目安となる適当な階層までにとどめておくことが望ましい。

Step2

表 1 は、語義要素の階層 1 の全語義木中での出現回数（全語積中での出現回数）の順位が、階層が 2~10 の範囲で変わることによって変動する際の、変動の大きさの度合いを示している。すなわち、記号#は、順位が 100 位を超えて下がっていることを表し、記号*は、順位が 100 位を超えて上がっていることを表す。100 語程度の基本語を抽出する場合、#と*のつき方から、以下のことが分かる。

- 出現回数が多く、どの階層にも#や*がついていない語は、基本語の可能性が高い。
- #がついた語は、基本語でない。
- *がついた語は、基本語の可能性がある。

ここで、多くの場合、#または*の有無は、階層 5 ぐらいまでで決まる。特に出現回数の多い上位の語（たとえば、全語積中での出現回数が 20 回以上の語）については、若干の例外を除き、一度ついた記号は、階層が n から $n+1$ に変わっても、消えることはない。

また、2~10 の階層ごとの#または*がついている語義要素の個数は、表 2 のようになる。#や*がついている語義要素の個数は、階層 5 までは増加する傾向にあるが、階層 5 以降はほとんど変わらない。したがって、語義木の階層は、概ね 5 まででよいと考える。

Step 3

表 3 と表 4 は、階層 5~10 の語義木について、出現回数の多い上位 100 語を抽出したもので

1		2		3		4		5	
順位	語義要素	順位	語義要素	順位	語義要素	順位	語義要素	順位	語義要素
1	こと	1	こと	1	こと	1	こと	1	こと
2	する	2	もの	2	もの	2	もの	2	もの
3	もの	3	する	3	する	3	する	3	する
4	ようす	4	ようす	4	ようす	4	ようす	4	ようす
5	ひと	5	ところ	5	やる	5	いう	5	いう
6	ところ	6	ひと	6	なる	6	ことば	6	ことば
7	なる	7	なる	7	ところ	7	やる	7	おもう
8	つかう	8	やる	8	ことば	8	なる	8	ところ
9	とき	9	ある[有る]	9	ある[有る]	9	ある[有る]	9	なる
10	つくる	10	つかう	10	いう	10	ある[有る]	10	やる
11	どうく	11	とき	11	ひと	11	おもう	11	ぶぶん
12	ある[有る]	12	なか[中]	12	とき	12	ぶぶん	12	ある[有る]
13	おおきい	13	どうぶつ	13	ぶぶん	13	ちがう	13	かんじる
14	なか[中]	14	ことば	14	どうぶつ	14	ひと	14	ちがう
15	ちいさい	15	からだ	15	ちがう	15	かたち	15	かたち
16	どうぶつ	16	つくる	16	なか[中]	16	まとめる	16	ひと
17	いれる	17	うえ[上]	17	まとめる	17	かんじる	17	ばしょ
18	みず	18	みず	18	かたち	18	ばしょ	18	め[目]
19	うえ[上]	19	いう	19	つかう	19	どうぶつ	19	まとめる
20	たべもの	20	ぶぶん	20	ばしょ	20	なか[中]	20	おなじ
21	すむ[住む 棲む]	21	かたち	21	かんじる	21	しよくぶつ	21	なか[中]
22	き[木]	22	しよくぶつ	22	おもう	22	とき	22	かく[書く 描く]
23	だす	23	おおきい	23	うえ[上]	23	おなじ	23	あらわす
24	おと	24	たべもの	24	しよくぶつ	24	うつす[移す]	24	はえる[生える]
25	からだ	25	つける[付ける 着ける]	25	うつす[移す]	25	き[木]	25	しよくぶつ
26	かたち	26	そと	26	からだ	26	うえ[上]	26	もつ[持つ]
27	たべる	27	ちいさい	27	き[木]	27	もつ[持つ]	27	はなす[話す]
28	たくさん	28	うつす[移す]	28	くらべる	28	め[目]	28	つたえる
29	つける[付ける 着ける]	29	かんじる	29	うごかす	29	かく[書く 描く]	29	うえ[上]
30	とり	30	き[木]	30	わたしたち	30	はえる[生える]	30	どうぶつ
31	ない	31	わたしたち	31	そと	31	そと	31	うつす[移す]
32	でる	32	いれる	32	もつ[持つ]	32	あらわす	32	そと
33	よい	33	くらべる	33	みえる	33	みえる	33	あたま
34	つく[つく]	34	うごかす	34	たべもの	34	その	34	き[木]
35	した[下]	35	たべる	35	あたま	35	うごかす	35	みえる
36	きかい[機械]	36	ない	36	あし[足]	36	つたえる	36	その
37	いう	37	だす	37	その	37	くらべる	37	とき
37	うみ[海]	38	あし[足]	38	ない	38	からだ	38	ただし
37	さかな	38	ちから	39	おおきい	39	あたま	39	かわる[変わる]
40	ながい	40	まとめる	40	め[目]	40	わたしたち	40	かんがえる
40	はな[花]	41	どうく	41	かく[書く 描く]	41	はなす[話す]	41	からだ
42	いろいろ	42	たくさん	42	おと	42	おと	42	うごかす
42	て[手]	42	ばしょ	43	くさ	43	くさ	43	わたしたち
42	なに	43	て[手]	44	みず	44	ある	44	くらべる
42	やさしい	45	おと	45	だす	45	たべもの	45	かお
46	ちから	46	もつ[持つ]	46	ちいさい	46	つかう	46	おと
47	え[顔]	47	おもう	47	つくる	47	あし[足]	47	みる
47	とても	48	みえる	48	はえる[生える]	48	ただし	48	て[手]
49	あし[足]	49	あたま	49	やく	49	だす	49	だす
49	ことば	50	ながい	50	ある	50	おく[置く]	50	かっこう
49	そら	51	たかい[高い]	50	おなじ	51	ばあい	51	いっしょ
52	いえ	51	やく	52	たてる[立てる]	52	かお	52	おく[置く]
52	いきもの	54	ちがう	53	ばあい	53	はなれる	53	でる
52	のりもの	54	たてる[立てる]	54	たべる	54	て[手]	54	たべもの
52	ふく[服]	55	すむ[住む 棲む]	55	いける	55	かわる[変わる]	55	あし[足]
56	かく[書く 描く]	55	その	55	て[手]	56	でる	56	くさ
57	あそび	55	はえる[生える]	55	でる	57	うつる[移る]	57	とる[取る]
57	こども	58	め[目]	58	つける[付ける 着ける]	58	ながい	58	こえ
57	むし[虫]	59	ほう	59	いれる	59	おおきい	59	ある
60	おみせ	60	かく[書く 描く]	60	ながい	60	たべる	60	ほう
60	おもう	61	でる	61	あらわす	61	たかい[高い]	61	うつる[移る]
60	じぶん	61	のりもの	61	たかい[高い]	62	うちがわ	62	え[顔]
60	ぬの	63	え[顔]	63	うちがわ	63	ない	63	はなれる
60	ほう	64	しごと	64	くだもの	64	みる	64	くち
65	きれい	65	とり	64	たくさん	65	ほう	65	じ[字]
65	じどうしゃ								

図5 語義木の階層の違いによる語義要素の出現回数と順位の変動

ある（表3は1～50位，表4は51～100位）。ここで，その語が階層5～10のいずれかで100位までに入っていない場合は，背景が灰色になっているが，そのような語は，各階層の100語中に17語しかない。すなわち，背景が灰色でない83語は，階層5～10のすべてにおいて100位までに入っている。そこで，これらの語を基本語として抽出することもできるが，この方式では，さらに深い階層まで調べてい

くと，抽出される語が減っていく可能性があり，階層をどこまで調べるかが問題になる。そこで，階層5で抽出された100語のうち80%以上が，より深い階層においても抽出されることから，本稿では，階層5の語義木における上位100語を基本語として抽出することにする。

表3 基本語の候補 (1~50位)

5		6		7		8		9		10	
順位	語義要素	順位	語義要素	順位	語義要素	順位	語義要素	順位	語義要素	順位	語義要素
1	こと	1	こと	1	こと	1	こと	1	こと	1	こと
2	もの	2	もの	2	もの	2	もの	2	もの	2	もの
3	する	3	する	3	する	3	する	3	する	3	する
4	ようす	4	ようす	4	ようす	4	ようす	4	ようす	4	ようす
5	いう	5	いう	5	ぶぶん	5	ぶぶん	5	ぶぶん	5	ぶぶん
6	ことば	6	ことば	6	いう	6	いう	6	かんじる	6	かんじる
7	おもう	7	おもう	7	おもう	7	かんじる	7	いう	7	ある[有る]
8	ところ	8	ぶぶん	8	ことば	8	おもう	8	ある[有る]	8	いう
9	なる	9	なる	9	やる	9	ことば	9	おもう	9	おもう
10	やる	10	やる	10	かんじる	10	やる	10	ことば	10	ことば
11	ぶぶん	11	かんじる	11	なる	11	ある[有る]	11	やる	11	やる
12	ある[有る]	12	ところ	12	ある[有る]	12	なる	12	なる	12	なる
13	かんじる	13	ある[有る]	13	ところ	13	め[目]	13	め[目]	13	め[目]
14	ちがう	14	ちがう	14	かたち	14	かたち	14	かたち	14	かたち
15	かたち	15	かたち	15	め[目]	15	ところ	15	ところ	15	ところ
16	ひと	16	め[目]	16	ちがう	16	ちがう	16	ちがう	16	はえる[生える]
17	ばしょ	17	ひと	17	なか[中]	17	はえる[生える]	17	はえる[生える]	17	ちがう
18	め[目]	18	なか[中]	18	ひと	18	あたま	18	あたま	18	あたま
19	まとめる	19	あらかず	19	あたま	19	なか[中]	19	なか[中]	19	なか[中]
20	おなじ	20	おなじ	20	はえる[生える]	20	ひと	20	ひと	20	ひと
21	なか[中]	21	かく[書く描く]	21	あらかず	21	あらかず	21	あらかず	21	みる
22	かく[書く描く]	22	はえる[生える]	22	はなす[話す]	22	みる	22	みる	22	もつ[持つ]
23	あらかず	23	ばしょ	23	かく[書く描く]	23	もつ[持つ]	23	もつ[持つ]	23	そと
24	はえる[生える]	24	はなす[話す]	24	おなじ	24	はなす[話す]	24	そと	24	あらかず
25	しよくぶつ	25	あたま	25	もつ[持つ]	25	そと	25	うえ[上]	25	かお
26	もつ[持つ]	26	まとめる	26	みる	26	かく[書く描く]	26	はなす[話す]	26	うえ[上]
27	はなす[話す]	27	もつ[持つ]	27	そと	27	うえ[上]	27	かく[書く描く]	27	はなす[話す]
28	つたえる	28	そと	28	ばしょ	28	おなじ	28	かお	28	かく[書く描く]
29	うえ[上]	29	しよくぶつ	29	まとめる	29	ばしょ	29	おなじ	28	その
30	どうぶつ	30	かんがえる	30	うえ[上]	30	まとめる	30	その	30	かっこう
31	うつす[移す]	31	うえ[上]	31	かんがえる	31	かお	31	かっこう	31	おなじ
32	そと	32	つたえる	32	どうぶつ	32	どうぶつ	32	ばしょ	32	ばしょ
33	あたま	33	どうぶつ	33	しよくぶつ	33	かんがえる	33	まとめる	33	どうぶつ
34	き[木]	34	かわる[変わる]	34	かお	34	その	34	どうぶつ	34	みえる
35	みえる	35	みる	35	つたえる	35	かっこう	35	かんがえる	35	まとめる
36	その	36	うつす[移す]	36	かわる[変わる]	36	つたえる	36	みえる	36	かんがえる
37	とき	37	その	37	その	37	しよくぶつ	37	つたえる	37	からだ
38	ただしい	38	ただしい	38	うつす[移す]	38	かわる[変わる]	38	からだ	38	つたえる
39	かわる[変わる]	39	みえる	39	からだ	39	みえる	39	かわる[変わる]	39	おと
40	かんがえる	40	かお	40	かっこう	40	からだ	40	しよくぶつ	40	かわる[変わる]
41	からだ	41	からだ	41	ただしい	41	うつす[移す]	41	おと	41	でる
42	うごかす	42	とき	42	みえる	42	ただしい	42	でる	42	ながい
43	わたしたち	43	かっこう	43	て[手]	43	おと	43	うつす[移す]	43	はな[鼻]
44	くらべる	44	き[木]	44	おと	44	でる	44	て[手]	44	うつす[移す]
45	かお	45	て[手]	45	え[絵]	45	て[手]	45	のびる	45	て[手]
46	おと	46	いっしょ	46	じ[字]	46	え[絵]	46	はな[鼻]	46	のびる
47	みる	47	おと	47	いっしょ	47	じ[字]	47	ただしい	47	しよくぶつ
48	て[手]	48	わたしたち	48	とき	48	わたしたち	48	ながい	48	だす
49	だす	49	だす	49	わたしたち	49	だす	49	だす	49	こえ
50	かっこう	50	え[絵]	50	でる	50	いっしょ	50	こえ	50	とき

表4 基本語の候補 (51~100位)

5		6		7		8		9		10	
順位	語義要素	順位	語義要素	順位	語義要素	順位	語義要素	順位	語義要素	順位	語義要素
51	いっしょ	51	でる	51	き[木]	51	とき	51	え[絵]	51	ただしい
52	おく[置く]	52	じ[字]	52	だす	52	うちがわ	52	じ[字]	52	わたしたち
53	でる	53	うごかす	53	いろ	53	のびる	53	とき	53	うちがわ
54	たべもの	54	くらべる	54	うちがわ	54	いろ	54	うちがわ	54	え[絵]
55	あし[足]	55	おく[置く]	55	こえ	55	はな[鼻]	55	わたしたち	55	いろ
56	くさ	56	こえ	56	のびる	56	こえ	56	いろ	56	じ[字]
57	とる[取る]	57	とる[取る]	57	おく[置く]	57	ながい	57	いっしょ	57	くち
58	こえ	58	のびる	58	うごかす	58	き[木]	58	くち	58	おもて[表]
59	ある	59	あし[足]	59	くらべる	59	くち	59	くらべる	59	くらべる
60	ほう	60	うちがわ	60	ながい	60	あし[足]	60	おもて[表]	60	いっしょ
61	うつる[移る]	61	しるす	61	あし[足]	61	おもて[表]	61	あし[足]	61	あし[足]
62	え[絵]	62	ほう	62	しるす	62	くらべる	62	き[木]	62	ほう
63	はなれる	63	くち	63	とる[取る]	63	しるす	63	ほう	63	いきる
64	くち	64	うつる[移る]	64	はな[鼻]	64	おく[置く]	64	かみのけ	64	かみのけ
65	じ[字]	65	いきる	65	くち	65	うごかす	65	くび	65	くび
66	うちがわ	66	たべもの	66	おもて[表]	66	いきる	66	いきる	66	とる[取る]
67	ばあい	67	くさ	67	いきる	67	かみのけ	67	しるす	67	しるす
68	のびる	68	ながい	68	ほう	68	くび	68	うごかす	68	ぜんぶ
69	たべる	69	いろ	69	うつる[移る]	69	ほう	69	とる[取る]	69	うごかす
70	いきる	70	おもて[表]	70	いろいろ	70	とる[取る]	70	おく[置く]	70	うつる[移る]
71	たかい[高い]	71	はなれる	71	ぜんぶ	71	いろいろ	71	ぜんぶ	71	たくさん
72	ながい	72	ぜんぶ	72	しゃべる	72	ぜんぶ	72	うつる[移る]	72	おく[置く]
73	みず	73	みず	73	かみのけ	73	うつる[移る]	73	たくさん	73	き[木]
74	ぜんぶ	74	たべる	74	くび	74	たかい[高い]	74	いろいろ	74	いろいろ
75	おもて[表]	75	しゃべる	75	みず	75	たくさん	75	たかい[高い]	75	たかい[高い]
76	いれる	76	ある	76	くさ	76	しゃべる	76	いきもの	76	いきもの
77	しるす	77	はな[鼻]	77	さき	77	さき	77	みず	77	うつる[映る]
78	おおきい	78	たかい[高い]	78	たべもの	78	みず	78	しゃべる	78	みず
79	いきもの	79	いきもの	79	よい	79	いきもの	79	さき	79	さき
80	ちいさい	80	ばあい	80	たかい[高い]	80	よい	80	うつる[映る]	80	しゃべる
81	わかる	81	うつる[映る]	81	いきもの	81	きちんと	81	みみ	81	みみ
82	うつる[映る]	82	さき	82	たべる	82	まちがう	82	たべる	82	おおきい
82	のせる	83	きちんと	83	きちんと	83	くさ	83	きちんと	83	たべる
84	さき	84	まちがう	84	まちがう	84	みみ	84	まちがう	84	かた
85	いろ	85	いろいろ	85	ある	85	たべる	85	よい	85	てくび
86	しらべる	86	わかる	86	うつる[映る]	86	うつる[映る]	86	おおきい	86	きちんと
86	つかう	87	よい	87	わかる	87	たべもの	87	かんじ	87	まちがう
88	ない	88	かみのけ	88	ばあい	88	かた	88	くさ	88	かんじ
89	むし[虫]	89	くび	89	たくさん	89	てくび	89	かた	89	ちいさい
90	はな[鼻]	90	ちいさい	90	はなれる	90	かきあらわす	90	てくび	90	よい
91	ちから	91	のせる	91	みみ	91	ある	91	かきあらわす	91	かきあらわす
92	かむ	92	いれる	92	かた	92	ひとつ[1つ]	92	きごう	92	ばあい
93	よこ	93	おおきい	93	てくび	93	いちど	93	ちいさい	93	きごう
94	くだもの	94	むし[虫]	94	ひとつ[1つ]	94	おおきい	94	たべもの	94	におい
95	うごく	95	みみ	95	いちど	94	わかる	95	ひとつ[1つ]	95	きく[聞く]
96	かみのけ	96	しらべる	96	おおきい	96	きごう	96	いちど	96	あか[赤]
96	みみ	97	ない	97	ちいさい	97	かんじ	97	ばあい	97	うける
98	くび	98	かた	98	かきあらわす	98	ばあい	98	きく[聞く]	98	あお
99	しゃべる	99	てくび	99	むし[虫]	99	ちいさい	99	わかる	99	ひとつ[1つ]
100	たくさん	100	きく[聞く]	100	きごう	100	きく[聞く]	100	あか[赤]	100	いちど
								100	うける		

6. 考察

(1) 語積がない基本語

前章で抽出した 100 語の基本語の候補のうち、本辞典において語積がないものは、以下の 18 語である。

あらわす, ある, うちがわ, かっこう,
かんじる, こと, しるす, その, つたえる,
ところ, なる, ばあい, ぶぶん, ほう,
まとめる, もの, やる, ようす

語積がない理由としては、以下が推測される。

- イラストに表しにくい。
- 抽象度の高い概念である。
- 語積においては必要であるが、幼児にとってあまり身近ではない概念である。

しかし、下線を引いた 8 語は、必ずしもこの理由に該当しないと思われるので、語積を記載するという選択もありえると考ええる。

(2) 見出し語の選定に関する辞書の特徴

全語積中での出現回数が多い語と、全語義木中での出現回数が多い語を比較することによって、『三省堂 こどもことば絵じてん』における見出し語の選定上の特徴を考える。

全語積中での出現回数が多い語を抽出することによって、それらの語に関連する語が、見出し語として多く収録されていることが分かる。これらの全語積中での出現回数が多い（基本的な）語を除外したものは、当該辞書の見出し語の選定上の特徴を示していると考えられる。なぜならば、当該辞書には、それらの基本的でない（特徴的な）語に関連する語が見出し語として多く収録されていることになるからである。このような特徴的な語は、特に同種の辞書を比較する場合に参考になる。

表 1 において全語積中での出現回数が 20 回以上と多いにも関わらず、100 語の基本語で

はないものは、以下の 42 語である。

あじ[味], あそび, いい, いえ, いく,
いろいろ, うみ[海], うる[売る], おみせ,
おんな, かぜ[風], がっき, きかい[機械],
きる[着る], きれい, こども, さかな,
した[下], じどうしゃ, じぶん, すば一つ,
すむ[住む_棲む], そら, つく[付く],
つくる, つける[付ける_着ける], つち,
どうぐ, とても, とり, とる[取る], なに,
ぬの, のりもの, はな[花], ひかり, ふく[服],
ほか, まえ, やさい, よい, りょうり

これらのうち下線を引いた 22 語は、幼児にとって身近な概念であると思われる。本辞典には、そのような概念に関連する見出し語が多く収録されており、この点が、幼児向けの国語辞典らしい特徴となっているのではないか。

7. 関連研究

(1) 辞書学と辞書批評

辞書学とは、辞書、および、辞書の編集方法を研究対象とする学問領域である。辞書学の成果のひとつとして、辞書批評がある。商業出版物である辞書を詳細に調べて、厳格な批評を行うもので、通例、書評の形をとる。その目的のひとつは、書評を通じて、当該辞書がどのようなものであるかを、読者に知らせることである。さらに、学術的な専門知識に基づいて批評が行われるため、辞書の改善に貢献すると考えられる[Jackson(2002)]。批評の対象となる項目は多岐にわたる。見出し語の選定、語積、用例の妥当性などの内容面だけでなく、検索の容易性やページレイアウトなどの形式面も対象となる。本稿では、見出し語の選定について、『三省堂 こどもことば絵じてん』の批評を行ったことになる。

(2) Thorndike の分類システム

英語の辞書において、見出し語をバランス

よく辞書に収録することを目的とする方法として、Thorndikeによって提案された分類システムがある[Landau(1984)]。この方法では、単語の語頭の綴りに基づく105のブロックから構成される表を利用する。すなわち、第1のブロックには、語頭がaの語からadkの語が入る、第2のブロックには、語頭がadlの語からalhの語が入る、・・・というようにブロックを定義する。各ブロックに入る語の綴りの範囲は、その範囲に該当するすべての英単語の個数がほぼ同じになることを想定して設計されている。たとえば、aで始まる語のブロックは6個であるが、cで始まる語のブロックは10個もあり、jで始まる語のブロックは1個しかない。ここで、辞書に収録する語の候補を、その綴りにしたがって各ブロックに分類していく。その結果、各ブロックに入る語の個数がほぼ同じになれば、見出し語がバランスよく収録されたと考える。この方法では、語の意味を考慮せずに、綴りだけに注目するが、専門用語辞典のような特化した目的の辞書ではなく、一般の辞書において見出し語のバランスを保つ上では有用であると思われる。

(3) 基本語彙の種類

2章でも述べたように、ある語が基本語であるかどうかを判定するために、いくつかの基準が考えられる。言い換えれば、基本語という概念には、複数の定義がありえる。林は、これを以下のように整理する[林(1971)]。

基礎語彙

一つの言語体系にもとづき、現実の言語体系に手を加えて作った半人工的モデル機構である。その範囲内で、普通のこととは一応何でも言い表わせるはずの、一つの完結体であり、小宇宙である。

基本語彙

「何かの事を行うために、第一段階でまず必要とされる語彙」という意味で使

うことにしたい。従って、「何かの事」が何であるかによって、基本語彙の性格も内容も規定されることになる。

基準語彙

その社会での特殊な用や、さらに細分化した社会内での用を弁ずることはできないが、一般的な生活は不自由なくできる—というために必要な語彙が、その社会での基準語彙である。

基調語彙

源氏物語の語彙調査をした寿岳章子氏は、見出された語彙と作品との関係を考えて、テーマ語彙と名づける一群の語を指定している。これなどに、基調語彙の好例を見ることができる。

基幹語彙

ある語集団の中に、その集団の骨格のような部分として、その集団をささえる基幹部分として、現に存在する、語の部分集団を呼ぶ。

本稿で扱った基本語は、『こどもことば絵じてん』の語積を理解することを目的としたものであり、基本語彙に相当すると考えられる。

(4) Basic English と基礎日本語

Ogden が考案した Basic English は、850語で日常のほとんどあらゆる分野にわたり、明快で英語らしい英語が書けるように工夫された補助的国際言語である。使用する単語を850語に限定しただけでなく、文法も英文法を簡略化したものになっている。したがって、通常の英語とは異なる半人工的な言語であるとみなせるため、Ogden が選定した850語は、林の定義における基礎語彙に相当すると考えられる。しかし、外国人などが英語を習得する際の第一段階として学ぶべき語彙であると捉えれば、英語教育を目的とした基本語彙で

あるともいえる。

Ogden は、20,000 語以上の英単語の語釈を Basic English の 850 語だけで記述した辞書を出版している[オグデン(1960)]。たとえば、見出し語「barber」の語釈は、「One whose trade is cutting and dressing hair on face and head」となっているが、この中で使用されている「one」、「who」、「trade」、「be」、「cut」、「and」、「dress」、「hair」、「on」、「face」、「head」は、いずれも Basic English として選定された語である。すなわち、すべての語義要素が Ogden の考える基本語になっているといえる。この辞書の見出し語について、本研究の語義木に相当するものを生成すれば、Ogden の基本語とは異なる意味での基本語が抽出できる可能性がある。

上居によって考案された基礎日本語[上居(1933)]は、Basic English の日本語版といえるもので、語彙として 1,000 語あまりが選定されている。たとえば、「体」は、「からだ」ではなく「たい」、「頭」は、「あたま」ではなく「かしら」として選定しているが、その理由は、これら 4 語の意味は、以下の通りであるため、「たい」と「かしら」の方が、「応用の範囲」と「言葉の経済」の点で基礎になるとしている。

- からだ —
人の肉体
- たい —
からだ
天体, 円体, 液体, 気体, 大體, 文体
- あたま —
頭 (動物の体の顔より上の部分)
考える力 (=脳)
- かしら —
頭
人々の上に立つ人
順序の第一 → 一例, 頭文字

(5) 語釈の循環に関する実験

水谷は、本研究の語義木と類似した方法で、

語釈の循環に関する解析を試みている[水谷(2012)]。語釈の循環とは、辞書で語 A を引くと、その語釈の中で語 B と語 C が使われていて、語 C の意味が分からなかったのので、語 C を引くと、その語釈の中で語 A や語 B が使われているような状況である。このような循環はすべて悪い訳ではなくて、辞書では必ずどこかで循環せざるを得ないとした上で、循環の様子を図に示すことによって、理にかなった循環であるかどうかを見極めようとした。このときの図が、語義木に相当すると考えられる。さらに、基本語だと思われる語を 50 語ほど選定して、それだけを使って語釈を書こうとしたときに、どうしても新たに必要になる語を基本語に追加する、・・・という作業を繰り返すことで、語釈の水準を高めていこうとする実験を試みている。これなどは、(4)の研究における Ogden の Basic English の辞書にも通じるものである。

(6) 語の基本度

水谷は、90 種類の雑誌で使用されている語の出現回数データのデータ[国立国語研究所(1962)]から、語の基本度を測定する関数を導き出し、この関数に基づいて、当該データの中から 700 語を基本語に選定した[水谷(2012)]。ただし、この基本語は、2 章で示した水谷による基本語の性質の 5) を主眼としたものである。基本度関数は、その語が使用されている雑誌の種類(評論・芸文, 庶民, 実用・通俗科学, 生活・婦人, 娯楽・趣味の 5 種に分類している)を考慮して、多方面にわたって使用されている語を抽出する。また、基本度関数を導く過程では、語が基本的である度合いを言語の専門家が主観的に判断しており、出現回数という客観的なデータのみから基本語を選定している訳ではない。本研究で提案した語義木を用いて、国語辞典に記載された語釈を分析することによって、より定量的に語の基本度が求められる可能性がある。

(7) 教育基本語彙の比較

日本語教育を目的とする基本語彙は、これまでもいくつか選定されてきた。国立国語研究所による『日本語教育基本語彙七種比較対照表』では、以下の7文献で提示された基本語彙を比較している[国立国語研究所(1982)]。

- (a) 日本語基本語彙 (岡本禹一)
- (b) 日本語教育における基礎学習語 (加藤彰彦)
- (c) 実用和英辞典 (海外技術者研修協会)
- (d) A CLASSIFIED LIST OF BASIC JAPANESE VOCABULARY (J.V.Neustupny)
- (e) 外国人のための基本語用例辞典 (文化庁)
- (f) 留学生教育のための基本語彙表 (樺島忠夫, 吉田弥寿夫)
- (g) 日本語教育基本語彙 第一次集計資料 (1) - 上位二千語 - (国立国語研究所)

比較の結果は、(g)に対する一致度、すなわち、(a)~(f)の各文献における基本語彙中の語で、(g)における基本語彙になっている語の割合としてまとめられている。一致度は、(a)は0.455、(b)は0.444、(c)は0.433、(d)は0.393、(e)は0.343、(f)は0.108となった。このように、一致度が高くないことから、基本語彙の選定は、非常に困難な作業であると考えられる。

(8) 教育基本語彙のデータベース

(7)の研究における基本語彙の比較は、2,000~4,000語程度を対象としているが、さらに大規模な語彙について調査した研究がある[国立国語研究所(2009)]。基本語彙の選定が非常に困難であることから、この研究では、選定した語を基本語彙として確定させるのではなく、基本語彙の候補となる語を網羅したデータベースを構築する、というアプローチをとる。このデータベースには、以下の7文献で提示された約27,000語が登録されている。

- (a) 教育基本語彙 (阪本一郎)
- (b) 新教育基本語彙 (阪本一郎)
- (c) 学習基本語彙 (田中久直)
- (d) 国語教育のための基本語体系 (池原檜雄)
- (e) 言語要素指導 (児童言語研究会)
- (f) 学習基本語彙 (中央教育研究所)
- (g) 日本語教育のための基本語彙調査 (国立国語研究所)

8. おわりに

本稿では、幼児向けの国語辞典に対して語義木を用いた解析を行うことで、本辞典の語彙における基本語を抽出した。それに基づいて、本辞典における見出し語の選定に関する考察を行った。しかしながら、現時点では、単一の辞書について実験的に試みた段階であり、その辞書の語彙も、幼児向けの簡単なものにとどまるため、本稿で示した手法の妥当性や有効性、特に、辞書編集において見出し語を選定する上での有効性を評価することは難しい。したがって、本手法を他の辞書に適用することが今後の課題である。

他の辞書としては、『三省堂 こどもことば絵じてん』と類似した幼児向けの国語辞典が考えられる。このような辞書は、三省堂以外にも出版されている[深谷(2015)][無藤ほか(2013)][篠崎(2008)]。それらに対して本手法を適用することによって、辞書間の比較が可能になる。その一方で、辞書間で共通する語が基本語として抽出される可能性もある。

さらに、一般の国語辞典や、日本語以外の言語の辞典(たとえば、英語の語の意味を英語で記した辞典、すなわち、英語を母国語とする人々にとっての国語辞典)について解析することによって、本手法の有効性を評価する必要があると考えている。特に後者では、日本語と英語における基本語の相違が明らかになるかもしれない。

また、専門用語辞典の類を解析することに

よって、当該辞典での基本語に相当する語を抽出することができる。そのようにして抽出された語が、真に当該専門領域における基本語となっているかどうかをチェックすれば、当該辞典における見出し語の選定の妥当性の評価に繋がると考えられる。

こうした課題に取り組むには、正確な語義データベースを効率的に構築できなければならない。本稿で示した解析を行う際に、最も労力を要したのは語義データベースの構築であった。紙の辞書を対象とする場合は、まず、語釈をコンピュータに入力する必要がある。次に、語釈から語を抽出していく。このとき、形態素解析ソフトウェアを活用することで、ある程度の作業の効率化が図れるが、最終的には、形態素解析結果を人間が確認、修正する必要がある。したがって、一般の国語辞典について、語義データベースを構築するには、多大な労力がかかるので、可能な限り効率的に作業を進める工夫が求められる。

参考文献

- [1] C.K.オグデン (監修) (1960)『The General Basic English Dictionary』北星堂書店。
- [2] 上居光知 (1933)『基礎日本語』六星社。
- [3] 深谷圭助 (監修) (2015)『角川 ことばえじてん』角川学芸出版。
- [4] 林四郎 (1971)「語い調査と基本語彙」『電子計算機による国語研究 III』pp.1-35 秀英出版。
- [5] 飯間浩明 (2013)『辞書を編む』光文社。
- [6] Jackson, H.(2002) *Lexicography: An Introduction*, Routledge.
- [7] 金田一春彦 (監修) (2009)『三省堂 ことばもことば絵じてん』三省堂。
- [8] 国立国語研究所 (1962)『現代雑誌九十種の用語用字 第一分冊 総記および語彙表』秀英出版。
- [9] 国立国語研究所 (1982)『日本語教育基本語彙七種比較対照表』大蔵省印刷局。
- [10] 国立国語研究所 (2009)『教育基本語彙の

基本的研究』明治書院。

- [11] 倉島節尚 (2008)『日本語辞書学への序章』大正大学出版会。
- [12] Landau, S.I.(1984) *Dictionaries: The Art and Craft of Lexicography*, Charles Scribner's Sons.
- [13] 増井元 (2013)『辞書の仕事』岩波書店。
- [14] MeCab 公式サイト「MeCab: Yet Another Part-of-Speech and Morphological Analyzer」<http://taku910.github.io/mecab/> (2017年2月1日)。
- [15] 水谷静夫 (1983)『朝倉日本語新講座 2 語彙』朝倉書店。
- [16] 水谷静夫 (2012)「理想の辞書を求めて」『ユリイカ』第44巻第3号 pp.64-69 青土社。
- [17] 無藤隆・天野成昭・宮田 Susanne (監修) (2013)『学研 ことばえじてん』学研マーケティング。
- [18] 篠崎晃 (監修) (2008)『小学館 ことばのえじてん』小学館。

謝辞

語義データベースへのデータ登録にご助力いただいた高橋裕幸氏に感謝いたします。