

新しい日本語辞書定義文型の策定に向けて

佐藤 理史 (名古屋大学大学院工学研究科)

夏目 和子 (名古屋大学大学院工学研究科)

Towards Full-Sentence Definitions of Japanese Words

Satoshi Sato (Graduate School of Engineering, Nagoya University)

Kazuko Natsume (Graduate School of Engineering, Nagoya University)

1 はじめに

我々は、語の意味や用法を調べるために、国語辞典を引く。これは、人間による、辞書の典型的な利用法である。

一方、計算機が辞書を利用する場合もある。我々のグループでは、クロスワードパズルのソルバー [1] や大学入試センター試験『国語』現代文のソルバー [2] などを作成しているが、そこでは、しばしば、ある表現の言い換えを取得したり、2つの言語表現の意味的等価性を判定したりすることが必要となる。国語辞書を利用してこれらを実行するプログラムを書くと、通常の利用ではなかなか気づきにくかった問題が見えてくる。語の定義に関して言えば、現存する国語辞典の多くは、次のような問題を抱えている。

- 辞書の使用者として、母語話者が想定されており、語の定義を理解するために、かなりの日本語力と常識が必要となる。(そのため、計算機は、国語辞書を十分活用できない。)
- 語の定義として、言い換え語が提示されるだけの場合があり、意味や用法に関する十分な情報が得られない。
- 語の定義(言い換え語)が、しばしば循環する。

これらの問題を一言で要約すれば、それは、「語の定義の方法が十分に統制されていない」ということに尽きる。

これに対して、英語の学習用の辞書では、20世紀後半から、非母語話者の辞書利用を容易にするための各種の工夫が導入されてきている。語の定義の統制に関しては、次の2つが導入されている。

- 語を定義するために使用する語彙の統制 (LDOCE [3] など)
- 完全な文 (full sentence) による、語の定義 (COBUILD [4])

日本語辞書に対しても、同じようなことができないであろうか。語の定義に使用する語彙の統制と文型の統制は、非母語話者にとって有益であるに留まらず、計算機が辞書を知識源として利用する際にも好都合である。

このような考えに基づき、我々は、日本語の語を定義する新しい方法の検討を開始した。今回、その一環として、約100語に対して実際に完全な文による定義 (full-sentence definition: FSD) を作成することを試みた。本稿では、その試みについて報告する。まず2節で、目指す方向性を示す。3節では、現時点の日本語 FSD 作成ガイドライン試案を示す。4節では、今回の試行で作成した FSD の具体例を示し、最後に5節で現状をまとめる。

2 目指す方向性

現在の国語辞書では、「旧態」を次のように説明する。

- 【旧態】 古くからの状態・ありさま。「一依然」(岩波国語辞典第7版)
 【旧態】 以前からの状態。また、昔のままの姿。「一を脱する」「一依然(=もとのままで少しも変化や進歩がないさま)とした経営方針」(明鏡国語辞典第2版)
 【旧態】 昔の状態。「一を存する／一依然(=元通りで、少しも変化・進展が無い)」(新明解国語辞典第6版)
 【旧態】 以前(から)のありさま。「一依然・一を脱する」(三省堂国語辞典第7版)

我々は、このような記述ではなく、おおよそ次のような定義が好ましいと考える。

- 【旧態】 《旧態依然 {の | とした | たる}》〈何か：制度・やり方〉とは、〈何か〉が昔から {変わっていない | 進歩していない} ことを [批判的に] いう。

この記述のポイントは、次の4点である。

1. 主に、「旧態依然」という形式で使用されることを示す。
2. 「旧態依然」が、典型的には「制度」「やり方」などを修飾することを示す。
3. 「旧態依然」と被修飾語の間には、「の」「とした」「たる」が使用できることを示す。
4. 「旧態依然」が「批判的」な意味を含意することを示す。

このように、定義の記述形式を工夫すれば、その語がどのように使用されるかについて、より豊富な情報を提供することができる。我々が設計したいのは、まさに、このような記述形式である。

3 日本語 FSD 作成ガイドライン試案

我々は、COBUILD の full-sentence definition (FSD) に倣い、語を完全な文で定義する方針を採用する。以下に、現時点でのガイドラインを示す。

3.1 対象とする語

当面、基本語以外の内容語を対象する。対象とする語の品詞は、名詞、動詞、形容詞(イ形容詞とナ形容詞)、副詞の4種類とする。

3.2 見出し語

(活用する) 見出し語の語形には、伝統的な形式、すなわち、動詞とイ形容詞は終止形、ナ形容詞は語幹を採用する。

3.3 ターゲット

一つの見出し語の下に、いくつかのターゲットを設定する。ターゲットとは、定義対象とする表現形式・用法・語義のことである。次のようなものをターゲットとする。

- | | | |
|-------------------|--------|----------|
| 1. 見出し語(それ自身) | 例：【語源】 | 《語源》 |
| 2. 見出し語の活用形・派生形 | 例：【巧妙】 | 《巧妙な》 |
| 3. 見出し語を用いた典型的な表現 | 例：【好評】 | 《好評を博する》 |

なお、本稿では、見出し語は【巧妙】のように【】で囲み、ターゲットは《巧妙な》のように《》で囲んで示す。

3.4 FSD の構成

FSD (full-sentence definition) を、前件部と後件部から構成する。以下に例を示す。

【合議】	(見出し語)
〈人々〉が〈あること〉について《合議する》とは、	(前件部)
〈あること〉について〈人々〉が互いに相談することをいう。	(後件部)

この例に示すように、「とは、」までが前件部、それ以降が後件部である。

3.5 前件部

前件部では、ターゲットの文型を示す。文型の示し方は、ターゲットの用法に依存する。以下では、主要な4用法(述語用法、連体修飾用法、連用修飾用法、名詞用法)について、典型的な方法を示す。

3.5.1 述語用法の前件部

ターゲットが述語用法の場合は、その述語の補足語(格要素と格助詞)を示すことで、その述語の文型を示す。以下に例を示す。

【心掛ける】	〈人〉が〈あること〉{を 〔をしよう〕に}を《心掛ける》とは、
【向上】	〈何か〉が《向上する》とは、
【好評】	〈何か〉が〔〈人々〉に〕《好評だ》とは、
【好評】	〈何か〉が〔〈人々〉に〕《好評を得る 博する》とは、

なお、格要素(〈人〉、〈あること〉、〈何か〉など)の記述方法は、後述する。

3.5.2 連体修飾用法の前件部

ターゲットが連体修飾用法の場合は、連体修飾用法の形式と典型的な被修飾語(体言)を示す。

【巧妙】	《巧妙な》〈方法・手段など〉とは、
【高尚】	《高尚 {な なる}》〈[人の] 精神・言動・作品・研究など〉とは、
【凝る】	《凝った》〈作品・料理など〉とは、

3.5.3 連用修飾用法の前件部

ターゲットが連用修飾用法の場合は、連用修飾用法の形式と典型的な被修飾語(用言)を示す。

【巧妙】	〔〈誰か〉が〈何か〉を〕《巧妙に》〈行なう〉とは、
【公然と】	〔〈人〉が〈あること〉を〕《公然と》〈行なう・主張する〉とは、

被修飾語となる用言(述語)を示すので、必然的に、その述語の補足語が必要となる。現時点では、暫定的にこれらの要素を記号[]で囲んで示すことにする。この記号は、付加的説明要素を表す。

3.5.4 名詞用法

ターゲットが名詞用法の場合は、その用法に典型的な修飾語が存在するか否かに応じて、異なる前件部の形式を採用する。

その用法に典型的な修飾語が存在する場合は、その修飾語を示す。典型的な修飾語が必要かどうかは、後件部を具体的にどう記述するかに密接に関わる。以下の例では、「〈ことば〉」

や「〈スポーツなど〉」への言及なしには、後件部を記述できない。特に、「コーチ」のように背後に動詞が隠されている名詞は、その動詞の格要素をノ格の修飾語としてとることができる[5]。

- 【語源】 〈ことば〉の《語源》とは、
その〈ことば〉のもとになった[別の]ことばのことをいう。
- 【コーチ】 [〈誰か〉の]〈スポーツなど〉の《コーチ》とは、
[〈誰か〉に]〈スポーツなど〉を教える人のこと。

一方、典型的な修飾語を持たない場合は、前件部をターゲットのみで構成する。

- 【暦】 《暦》とは、
一年の月・日・曜日に行事、祝日などが順に書かれたものをいう。

3.6 クラスまたは典型例の表示

格要素、典型的な被修飾語、修飾語は、記号〈〉で囲んで表示する。おおよそ、次のような方法がある。

1. 抽象度が高いクラスを提示する。
 - (a) クラス名を提示する。例：〈人〉、〈こと〉、〈もの〉、〈行ない〉
 - (b) 疑問表現で提示する。例：〈誰か〉、〈何か〉、〈どこか〉
 - (c) 不特定表現で提示する。例：〈ある人〉、〈あること〉、〈あるもの〉、〈ある行ない〉
 - (d) 代名詞で提示する。例：〈あなた〉、〈それ〉
 - (e) ラベルで表記する。例：〈A〉、〈B〉
 - (f) 複数であることを明示する。例：〈人々〉
2. 具体的なクラスを提示する。例：〈ことば〉、〈スポーツ〉、〈作品・料理など〉
3. 典型例を提示する。例：〈良い・目立っている・重要だ〉

さらに、これらを組み合わせる、末尾に「など」をつける、などの選択肢もある。

現時点では、これらの記述をどのように統制するかは定めず、FSDの具体例が十分に蓄積された段階で、記述方法の指針を定める予定である。

3.7 後件部

後件部の記述は、(概念的なレベルにおいては)次の2種類が存在する。

1. 言い換え的后件部：前件部と意味的に等価な言い換え表現(置換可能な表現)を提示する
2. 説明的後件部：前件部の意味を説明する表現を提示する。

実際の後件部は、これらのどちらかに排他的に分類されるということではなく、この2つの両極端の中間のどこかに位置付けられる。

語の意味を基本語のみで記述できると仮定することは、その意味を基本語の意味から合成できると仮定することである。たとえば、動詞「激怒する」は、基本となる動詞「怒る」にその程度が激しいことを組み合わせた語と考えることができる。これを図式的に表すと次のようになる。

ターゲット (述語用法)	程度を表す連用修飾語	基本となる述語
激怒する	激しく	怒る

上記のように、意味の合成が比較的単純な場合は、多くの文脈で置換可能な言い換え表現を考えることができ、後件部は、言い換え式的となる。しかし、意味合成が複雑になってくると、そのような言い換え表現を考えるのが難しくなり、後件部は説明的にならざるを得ない。

将来的には、典型的な意味合成のパターンを設定し、それぞれのパターンに対して、後件部の書き方を統制することが望ましい。しかしながら、現時点では、まずは、FSD の具体例を蓄積することを優先し、次に示すような最低限の原則のみを適用する。

3.7.1 原則

後件部の記述にあたっては、次のことを原則とする。

1. 前件部に現れた記号〈〉で囲まれた要素を含める。
2. 述語用法の場合は、その述語の動作性 (動作を表すか、状態を表すか)、意志性 (意志の有無) などの情報を含めるようにする。
3. 名詞用法の場合は、可能な限り、その名詞の上位語 (どのような基本語の下位に位置付けられるか) を示す。

3.7.2 言い換え式的後件部

後件部が言い換え式的であるということは、後件部自身が、そのまま普通に使える言語表現 (文型) であることを意味する。言い換え式的である場合は、後件部の最後を「こと」で終えることを原則とする。

- 【小柄】 《小柄な》〈人〉とは、
体が小さい〈人〉のこと。
- 【巧妙】 [〈誰か〉が〈何か〉を] 《巧妙に》〈行なう〉とは、
[〈誰か〉が〈何か〉を] {みごとに | うまく} 〈行なう〉こと。
- 【凝らす】 〈人〉が〈何か〉に 《目を凝らす》とは、
〈人〉が〈何か〉をよく見ること。
- 【根拠】 〈判断・主張・行動など〉の 《根拠》とは、
〈判断・主張・行動など〉のもとになる理由や基準のこと。

言い換え式的後件部では、特に、名詞用法と述語用法において、ターゲットの上位語を示すことが比較的容易である。上記の例では、「目を凝らす」の上位語は「見る」、「根拠」の上位語は「理由」または「基準」であることを明確に示している。

3.7.3 説明的后件部

ターゲットに対して適当な言い換え表現が存在しない場合は、その表現が意味するところを、説明的に示すしか方法がない。説明性が高い場合は、後件部の末尾に「ことをいう」を用いる。

- 【孤独】 《孤独な》〈人〉とは、
仲間や家族がなく、ひとりで寂しい〈人〉のことをいう。
- 【孤独】 〈誰か〉が《孤独を〈感じる〉》とは、
〈誰か〉がひとりで寂しいと〈感じる〉ことをいう。
- 【後悔】 〈ある人〉が〈ある行ない〉を《後悔する》とは、
〈ある人〉が〈ある行ない〉について、あとになって残念に思うことをいう。

3.7.4 使用する語彙

後件部は、基本語のみで記述するのが理想であるが、現時点では基本語の範囲を明確には定めていないので、そのような統制は採用しない。但し、ターゲット表現より平易な表現を用いることを心掛ける。

4 FSD 作成と具体例

今回、約 100 語に対して、実際に FSD を作成する試行を行なった。

1. 100 語は、旧日本語能力試験の出題基準 [6] の語彙表の「コ」で始まる語のうち、旧 1 級に位置付けられている語から選んだ。
2. FSD の作成に当たっては、既存の辞書 (主に、岩波国語辞典 (第 7 版新版)、明鏡国語辞典 (第 2 版)、三省堂例解小学国語辞典 (第 3 版)、三省堂国語辞典 (第 7 版)) と『現代日本語書き言葉均衡コーパス (BCCWJ)』の用例を参考にした。
3. FSD は、できるだけ旧 2 級以下の語彙で記述することとした。

以下に、品詞別に具体例を示す。なお、これらの具体例は、本稿執筆時点のものであり、完全に確定しているわけではない。ガイドライン試案の修正と並行して、日々修正を行なっている。

4.1 形容詞

形容詞は、一般に、述語用法、連体修飾用法、連用修飾用法をとる。それぞれの用法間で意味の変異が小さい場合は、ターゲットとして、頻度の高い用法を採用する。複数の語義や、表層形に差異がある場合は、複数のターゲットを立てる。以下に例を示す。

- 【固有】(1) 〈何か〉に《固有 {の | な}》〈性質・問題・知識など〉とは、
〈何か〉に特有な〈性質・問題・知識など〉のこと。
- (2) 〈何か〉に《固有の》〈財産・領土・権利など〉とは、
他から与えられたのではなく、もとからそこにあった〈財産・領土・権利など〉のことをいう。
- 【快い】(1) [〈誰か〉が〈何か〉を]《快く》〈思う・感じる〉とは、
[〈誰か〉が〈何か〉を] 楽しい・気持ちいいと〈思う・感じる〉こと。
- (2) [〈誰か〉が〈何か：依頼など〉を]《快く》〈行なう：引き受ける〉とは、
[〈誰か〉が〈何か〉を] 気持ちよく〈行なう〉ことをいう。
- (3) 〈風・音楽など〉が《快い》とは、
〈風・音楽など〉を気持ちよいと感じている様子をいう。

被連体修飾語や格要素の記述をどのように統制すべきかは、よく考える必要がある。英語の FSD では、後件部で “it” などの代名詞を使うのが自然であるが、日本語には馴染まないような気がする。〈財産・領土・権利など〉のように、典型的な例を 3 つ程度示したいが、そうすると後件部が長くなる。【快い】(2) で示したように、前件部では〈何か：依頼など〉のように〈クラス：具体例〉という形で示し、後件部では〈クラス〉で参照するという方法が、一つの候補かもしれない。

4.2 副詞

一般に、副詞の用法や意味を記述するのはかなり難しい。しかしながら、やっかいな副詞の大部分は基本語に含まれるので、現時点では対象となっていない。それでもなお、「殊に」

のような語が存在する。

【殊に】〈何か〉が《殊に》〈どうだ：良い・目立っている・重要だ〉とは
〈何か〉がとても〈どうだ〉ということ。

この例では、被連用修飾語の用言が形容詞相当の語であることを示そうとしている。

4.3 動詞

動詞では、どれだけの語義を認定し、独立したターゲットとして立てるかが問題となる(現時点では、それほど真剣に検討していない)。

【こじれる】(1) 〈話し合い・関係〉が《こじれる》とは、
[人と人との意見が合わず] 〈話し合い・関係〉が悪い方へ進んでしまうことをいう。

(2) 〈病気・問題〉が《こじれる》とは、
〈病気〉が悪くなったり、〈問題〉が解決しなかったりすることをいう。

【漉す】〈人〉が〈[液体状の] 何か〉を〈[ざるなどの] 道具〉で《こす》とは、
〈人〉が〈道具〉を使い、〈液体状のもの〉からそこに含まれる [小さな] かたまりを取り除くこと。

4.4 名詞

名詞は、おおきく、普通名詞、形容詞的な名詞、サ変名詞に分けて検討している。

普通名詞は、あまり工夫のしようがない。複合名詞を作りやすい名詞の場合は、ターゲットにそれを含める。

【行為】(1) 〈誰か〉の《行為》とは、
〈誰か〉の行ないのこと。

(2) [〈誰か〉の] 《〈犯罪・違反・危険〉行為》とは、
[〈誰か〉による] 〈犯罪・違反・危険〉を伴う行ないのことをいう。

形容詞的な名詞とは、品詞区分では名詞に分類されるが、意味的にはある種の状態や状況を表すものをさす。用法的には、やっかいなものが多い。

【交互】〈A〉と〈B〉を《交互に》〈行なう〉とは、
A B A Bのように、〈A〉と〈B〉を順に繰り返して〈行なう〉ことをいう。

【個別】(1) 《個別の》〈何か〉とは、
{ひとつひとつの | それぞれの} 〈何か〉のこと。

(2) 〈何か〉を《個別に》〈扱う：決める・知らせる・包むなど〉とは、
〈何か〉をそれぞれ別々に〈扱う〉こと。

サ変名詞とは、接尾辞「する」を伴って動詞化する名詞をさす。原則として述語用法をターゲットとするが、名詞用法もターゲットとすべき場合がある。

【合議】〈人々〉が〈あること〉について《合議する》とは、
〈人々〉が〈あること〉について互いに相談することをいう。

【講習】(1) 〈知識・技術など〉の《講習(会)》とは、
[希望する人々を集めて] 〈知識・技術など〉について教える会合のことをいう。

- (2) 〈誰か〉が〈知識・技術など〉について〔〈人々〉に〕《講習する》とは、
〔講習（会）に参加した〈人々〉に〕〈誰か〉が〈知識・技術など〉について
教えることをいう。
- (3) 〈誰か〉が〈知識・技術など〉について《講習を受ける》とは、
〈誰か〉が講習（会）に参加して、〈知識・技術など〉について学ぶことをい
う。

5 現状のまとめ

これまでの試行により、前件部の文型の示し方については、ほぼ固まってきたと考えている。主要な用法は4種類であり、ターゲットの用法を定めれば、示すべき文型はおおよそ決まってくる。しかしながら、たとえ動詞であっても、連体修飾用法を示すべき場合（「凝った」）もあり、品詞だけから示すべきターゲットの形式・用法を限定することはできない。それぞれの語に対して用例を観察し、どのような形式・用法をターゲットにすべきかを決定する必要がある。

格要素や被修飾語などの〈〉の中身をどのように選定・記述すべきかは、用例を観察しても悩む場合が多い。この問題に対しては、短期的には、より簡単に用例を観察することができるようツールを実現することで対処する予定である。しかしながら、長期的には、標準的に使用する語彙リスト（階層的シソーラス形式となっていることが望ましい）を、体言（名詞）と用言（動詞・形容詞）に対して定めることが必要になろう。

現時点であまり方針が固まっていないのは、後件部の文型および語彙の統制方法である。これらを統制するためには、意味合成の類型化が不可欠と考えているが、これをどのように進めるべきかについては、明確なアイデアはない。より多くのFSDを蓄積し、そこから徐々に指針が見えてくることを期待したい。

謝辞 本研究では、『現代日本語書き言葉均衡コーパス』を利用した。本研究は、JSPS 科学研究費基盤研究(B)「平易な日本語表現への工学的アプローチ」(課題番号 24300052)の助成を受けている。

参考文献

- [1] 内木賢吾, 佐藤理史, 駒谷和範. 日本語クロスワードを解く：性能向上の検討. 2013 年度人工知能学会全国大会 (第 27 回), 2D1-4, 2013.
- [2] 佐藤理史, 加納隼人, 西村翔平, 駒谷和範. センター試験『国語』現代文の傍線部問題を解くベースライン法. 情報処理学会研究報告, Vol. 2013-NL-212, No. 5, 2013.
- [3] Michael Mayer, editor. *Longman Dictionary of Contemporary English, 5th Edition*. Pearson Education Limited, 2009.
- [4] John Sinclair, editor. *COBUILD Advanced Dictionary of English, 7th Edition*. National Geographic Learning, 2012.
- [5] Sadao Kurohashi and Yasuyuki Sakai. Semantic analysis of japanese noun phrases - a new approach to dictionary-based understanding. In *Proc. of ACL-1999*, pp. 481–488, 1999.
- [6] 財団法人日本国際教育協会国際交流基金. 日本語能力試験出題基準【改訂版】. 凡人社, 2002.