

形態素解析に関する提案

—日本語教育の視点から—

山内博之

1. 問題意識

日本語教育において、語彙リストを作るということは、必要不可欠なことである。特に、読解教育においては、教科書を作る時にも、教材を作る時にも、語彙リストを付すことが非常に多い。語彙リストを作成するという作業は、ごく単純な作業だと思われがちであるが、実は、そう簡単なことではなく、かなり困難なことである。その困難さの1つに、語の単位の認定という問題がある。

語彙リストを作成する際、広く使用されているのが、「茶筌」¹⁾という形態素解析システムである。しかし、日本語教育という視点から見ると、この茶筌は、実は、それほど使いやすいものではない。その理由の筆頭が、先に述べた語の単位の認定の問題である。

次の(1)を見ていただきたい。

(1) 鈴木さんが来たんです。

この(1)を茶筌にかけると、次頁の表1のような解析結果を得ることができる。

日本語教育という視点で見た場合、ここで問題となるのが「んです」の扱いである。「鈴木さんが来たんです。」の「んです」は、表1を見ると、茶筌では「ん(名詞-非自立-一般)」と「です(助動詞)」の2語に分けられていることがわかる。しかし、日本語教育においては、「んです」は1語であると考えられている。

たとえば、現在、最も広く使用されている日本語教科書『みんなの日本語 初級Ⅱ 本冊』の第26課の練習Aにおいては、次頁の(2)のような図式を使用して、「んです」の説明が行なわれている。

(2)

表1. 「鈴木さんが来たんです。」の解析結果

表層語	読み	基本形	品詞	活用	活用形
鈴木	スズキ	鈴木	名詞-固有名詞-人名-姓		
さん	サン	さん	名詞-接尾-人名		
が	ガ	が	助詞-格助詞-一般		
来	キ	来る	動詞-自立	カ変・来ル	連用形
た	タ	た	助動詞	特殊・タ	基本形
ん	ン	ん	名詞-非自立-一般		
です	デス	です	助動詞	特殊・デス	基本形
。	。	。	記号-句点		

(2)

いく	んです	きれいな	んです
いかない		きれいじゃない	
いった		きれいだった	
いかなかった		きれいじゃなかった	

さむい	んです	びょうきな	んです
さむくない		びょうきじゃない	
さむかった		びょうきだった	
さむくなかった		びょうきじゃなかった	

この(2)では、動詞「いく」、形容詞「さむい」、形容動詞「きれいだ」²⁾、名詞+ダ「びょうきだ」に「んです」が付加される形になっており、「んです」を1語だと認識していることは明らかである。そして、当然のことながら、日本語教育の現場においても、「んです」は1語だとして、学習者に提示されることになる。

このように、語の単位の認定に関しては、茶筌と日本語教育の間に相違があることは明らかである。しかし、大量のデータを瞬時に解析することのできる茶筌のようなソフトは、日本語教育関係者にとっても非常に魅力的であり、何とか、いい形で使用できるようになってほしいものである。そのためには、ここで述べたような、語の単位の認定に関する問題をクリアしなければならない。本稿においては、どのように語の単位の認定を行えば、茶筌などのソフトが、日本語教育関係者にとってさらに使いやすいものになるのか、ということ考察する。

なお、コンピュータ解析を念頭において、日本語教育の視点から、語の単位の認定方法を考えるという試みは、管見の限りでは、いまだ行なわれたことがないのではないかと思われる。

2. 「のだ」について

前節では「んです」を例にとって、語の単位の認定に関する、茶筌と日本語教育の相違を指摘した。本節では、その問題意識を受け、どのように単位の認定を行えば「んです」をうまく扱えるようになるのか、ということを考える。

前節では、日本語教育においては「んです」は1語であると考えられているということを述べたが、では、「んです」が1語であるとする、(1)の文はどのように解析されるべきであるのか。表1を書き改めたものが、次の表2である。なお、簡略化のため、「読み」「活用」「活用形」の欄は省略した。

表1と表2の違いは、「んです」の扱いである。表1では、「ん」と「です」が分けて示されていたのだが、表2では「んです」が1語であると認定され、「助動詞」という品詞名が付されている。さらに、「んです」の基本形は「んです」ではなく「のだ」となっている。つまり、「んです」という助動詞が存在するのではなく、「のだ」という助動詞が存在し、「んです」はその変化形だということである。

表2. 「表1」の修正案

表層語	基本形	品詞
鈴木	鈴木	名詞－固有名詞－人名－姓
さん	さん	名詞－接尾－人名
が	が	助詞－格助詞－一般
来	来る	動詞－自立
た	た	助動詞
んです	のだ	助動詞
。	。	記号－句点

そうすると、次の(3)における「のだ」はもちろんのこと、(4)～(8)における「んだ」「のです」「のだった」「のではない」「んじゃない」も、基本形「のだ」の変化形だということになる。

- (3) 鈴木さんが来たのだ。
- (4) 鈴木さんが来たんだ。

(4)

- (5) 鈴木さんが来たのです。
- (6) 鈴木さんが来たのだった。
- (7) 鈴木さんが来たのではない。
- (8) 鈴木さんが来たんじゃない。

たとえば、(6)の解析結果を示すと表3のようになり、(7)の解析結果を示すと表4のようになるということである。

表3. 「のだった」の解析結果

表層語	基本形	品詞
鈴木	鈴木	名詞－固有名詞－人名－姓
さん	さん	名詞－接尾－人名
が	が	助詞－格助詞－一般
来	来る	動詞－自立
た	た	助動詞
のだった	のだ	助動詞
。	。	記号－句点

表4. 「のではない」の解析結果

表層語	基本形	品詞
鈴木	鈴木	名詞－固有名詞－人名－姓
さん	さん	名詞－接尾－人名
が	が	助詞－格助詞－一般
来	来る	動詞－自立
た	た	助動詞
のではない	のだ	助動詞
。	。	記号－句点

日本語教育においては、ある活用語の過去形・否定形などは、同語であるとして学習者に提出する場合が多い。つまり、「のではない」は、「のだ」に「は」と「ない」が付加されたものであるとは考えず、「のではない」全体で「のだ」の否定形だと考えることが多いということである。また、そう考えると、「んじゃない」を扱う際にも、「じゃ」の分割を考えずに済むという点で都合がよい。同様に、普通形と丁寧形も同語であると考えることが多い。しかし、以下の(9)(10)の例はどうであろうか。

- (9) 鈴木さんが来たのでもない。

(10) 鈴木さんが来たのでさえない。

日本語教育においては、「のだ」の否定形は「のではない」であり、「のでもない」「のでさえない」なども「のだ」の否定形であるとは、普通は考えないのではないか。したがって、「のでもない」は「ので (助動詞) + も (とりたて助詞) + ない (形容詞)³⁾」、「のでさえない」は「ので (助動詞) + さえ (とりたて助詞) + ない (形容詞)」と考えるのがよいのではないだろうか。

以上のことを鑑み、助動詞「のだ」の基本形と表層語(変化形)の関係を、次の(11)にまとめてみる。

(11) 助動詞「のだ」の基本形と表層語

方針：①「のだ」を「の+だ」とせず、1語であると認定する。

- ②「のです」の基本形を「のだ」とする。
- ③「のだった」の基本形も「のだ」とする。
- ④「のでした」の基本形も「のだ」とする。
- ⑤「のではない」の基本形も「のだ」とする。
- ⑥「のではありません」「のじゃない」「のではなく」等の基本形も「のだ」とする。
- ⑦「も」「さえ」など、「は」以外のとりたて助詞が現れた場合(例：のでもない)は、そのとりたて助詞の前で切れると考える(例：ので+も+ない〔助動詞+とりたて助詞+形容詞〕)。
- ⑧「んだ」「んです」は「のだ」「のです」と同等に扱う。
- ⑨「のか」「のだろう」「のでしょう」「のかもしれない」は2語であるとする。したがって、「か」「だろう」「でしょう」「かもしれない」等が接続している場合、表層語は「の」「ん」のみとなる。

基本形	表層語
「のだ」	《のだ》系 「のだ」「のだった」「ので」「ので(は)ない(です)」「のじゃない(です)」「ので(は)なかった(です)」「のじゃなかった(です)」「ので(は)ありません(でした)」「のじゃありません(でした)」「ので(は)なく(て)」「のじゃなく(て)」
	《のです》系 「のです」「のでした」

(6)

《んだ》系 「んだ」「んだった」「んで」「んで (は) ない (です)」「んじゃない (です)」「んで (は) なかった (です)」「じゃなかった (です)」「んで (は) ありません (でした)」「じゃありません (でした)」「んで (は) なく (て)」「じゃなく (て)」
《んです》系 「んです」「んでした」
《の・ん》のみ 「の」「ん」

かなり煩雑ではあるが、以上が「のだ」の基本形と表層語の関係である。日本語教育版の茶釜を作成するのなら、(11)の表の「表層語」のすべてを、「基本形」の「のだ」とリンクさせればよいということである。

基本形と表層語の関係が、助動詞「のだ」について明らかになったとすると、それを利用して助動詞「だ」についても、基本形と表層語の関係を記述することができる。助動詞「だ」の基本形と表層語の関係は、次の(12)のとおりである。

(12) 助動詞「だ」の基本形と表層語

方針：①「です」の基本形を「だ」とする。

②「だった」の基本形も「だ」とする。

③「でした」の基本形も「だ」とする。

④「ではない」の基本形も「だ」とする。

⑤「ではありません」「じゃない」「ではなく」等の基本形も「だ」とする。

⑥「も」「さえ」など、「は」以外のとりたて助詞が現れた場合（例：でもない）は、そのとりたて助詞の前で切れると考える（例：で+も+ない〔助動詞+とりたて助詞+形容詞〕）。

⑦「だろう」「でしょう」は、別の語として扱う。

基本形	表層語
だ	《だ》系 「だ」「だった」「で」「に」「で (は) ない (です)」「じゃない (です)」「で (は) なかった (です)」「じゃなかった (です)」「で (は) ありません (でした)」「じゃありません (でした)」「で (は) なく (て)」「じゃなく (て)」
	《です》系 「です」「でした」

基本形と表層語の関係が、助動詞「だ」について明らかになれば、同じ助動

詞の「ようだ」や「そうだ」における基本形と表層語の関係を示すことは、それほど難しいことではない。さらに、形容動詞の基本形と表層語の関係については、助動詞「だ」と基本的に同じである。次の(13)に、形容動詞における基本形と表層語の関係を「きれいだ」を例にして示しておく。

(13) 形容動詞「きれいだ」の基本形と表層語

基本形	表層語
きれいだ	《だ》系 「きれいだ」「きれいだっだ」「きれいで」「きれいに」「きれいで (は) ない (です)」「きれいじゃない (です)」「きれいで (は) なかった (です)」「きれいじゃなかった (です)」「きれいで (は) ありません (でした)」「きれいじゃありません (でした)」「きれいで (は) なく (て)」「きれいじゃなく (て)」
	《です》系 「きれいです」「きれいでした」

3. 動詞について

ここに表3を再掲するので、もう一度見ていただきたい。

表3. 「のだった」の解析結果

表層語	基本形	品詞
鈴木	鈴木	名詞－固有名詞－人名－姓
さん	さん	名詞－接尾－人名
が	が	助詞－格助詞－一般
来	来る	動詞－自立
た	た	助動詞
のだった	のだ	助動詞
。	。	記号－句点

表3を見ると、「のだった」が1語であると認定されているにもかかわらず、「来た」が2語に分けられていることがわかる。つまり、「のだ」においては、その過去形も同語であると認定されているのだが、「来る」においては、そうなるはないということである。これは明らかな矛盾であり、修正が必要である。では、どうすればよいのか。次頁の表5が、その修正案である。

表5では、「来た」を1語であると認め、その基本形を「来る」としている。つまり、「のだった」を1語であると認めて、その基本形を「のだ」としたの

(8)

表5. 「表3」の修正案

表層語	基本形	品詞
鈴木	鈴木	名詞－固有名詞－人名－姓
さん	さん	名詞－接尾－人名
が	が	助詞－格助詞－一般
来た	来る	動詞
のではない	のだ	助動詞
。	。	記号－句点

と、同様のことを行なったということである。

ここで、動詞においてル形とタ形を同語だと考えることが、日本語教育において妥当なことであるか否か、考えてみる。まず、次の(14)を見ていただきたい。(14)は、『みんなの日本語 初級 I 本冊』第4課の練習 A に掲載されている図である。

(14)

ね	ます	ね	ません	ね	ました	ね	ませんでした
やすみ	ます	やすみ	ません	やすみ	ました	やすみ	ませんでした
はたらき	ます	はたらき	ません	はたらき	ました	はたらき	ませんでした

この図を見ると、「ます」「ません」「ました」「ませんでした」が、動詞の語幹に同じように接続するという意味で同等に扱われていることがわかる。したがって、「ねます」「ねません」「ねました」「ねませんでした」の4者についても、同等に扱われていると考えて問題はないであろう。つまり、「ねます」「ねません」「ねました」「ねませんでした」は、同語の変化形であるというような扱いを受けているということである。

次に、(15)を見ていただきたい。(15)は、『みんなの日本語 初級 I 本冊』第20課の練習 A に掲載されている図の一部である。

(15)

普通形	丁寧形
かきます	かく
かきません	かかない
かきました	かいた
かきませんでした	かかなかった

ここには、「かきます」と「かく」、「かきません」と「かかない」、「かきました」と「かいた」、「かきませんでした」と「かかなかった」が、それぞれ、普通形と丁寧形の違いであることが示されている。つまり、これらのペアは、同語の変化形として扱われているということである。

(14)と(15)の2つの図から言えることは、たとえば「飲む」という動詞を例にとると、「飲みます」「飲みません」「飲みました」「飲みませんでした」「飲む」「飲まない」「飲んだ」「飲まなかった」の8つが同語だと認識されているということである。したがって、表5のように、「来た」を1語であると認定し、さらに、それを「来る」の変化形であるとみなすことは、日本語教育においては妥当なことであると思われる。

以上のことを鑑み、五段動詞「飲む」の基本形と表層語の関係をまとめたものが、次の(16)である。

(16) 五段動詞「飲む」の基本形と表層語

方針：①「飲まない」「飲みます」「飲んだ」「飲んで」などは基本形「飲む」の表層語であるとする。

②「飲まなければならない」は「飲ま(動詞) + なければならない(助動詞)」であるとする。

基本形	表層語
飲む	「飲む」「飲ま(ない・なかった)(です)」「飲み(ます・ました・ません・ませんでした)」「飲ん(で・だ)」「飲ま(ないで・なくて・なく・ず・ずに)」「飲め」「飲もう」「飲みなさい」

続いて、一段動詞「食べる」とサ変動詞「リラックスする」における基本形と表層語の関係を、(16)の「飲む」とほぼ同様の結果ではあるが、次の(17)(18)に示しておく。

(17) 一段動詞「食べる」の基本形と表層語

基本形	表層語
食べる	「食べる」「食べ(ない・なかった)(です)」「食べ(ます・ました・ません・ませんでした)」「食べ(て・た)」「食べ(ないで・なくて・なく・ず・ずに)」「食べろ」「食べよう」「食べなさい」

(10)

(18) サ変動詞「リラックスする」の基本形と表層語

基本形	表層語
リラックスする	「リラックスする」「リラックスし(ない・なかった)(です)」「リラックスし(ます・ました・ません・ませんでした)」「リラックスし(て・た)」「リラックスし(ないで・なくて・なく・ず・ずに)」「リラックスしろ」「リラックスしよう」「リラックスしなさい」

4. 形容詞について

述語の主要要素となる語の品詞名から日本語の文を分類すると、動詞文、形容詞文、形容動詞文、名詞文の4つになることは、よく知られていることである。つまり、述語を構成する活用語で重要なのは、動詞文・形容詞文・形容動詞文・名詞文それぞれの主要構成要素となる動詞・形容詞・形容動詞・助動詞「だ」である。本稿の第2節と第3節において、動詞・形容動詞・助動詞「だ」については、それぞれの基本形と表層語の関係を示したので、残るは形容詞のみである。次の(19)に、形容詞「多い」の基本形と表層語の関係を示す。

(19) 形容詞「多い」の基本形と表層語

方針：①「多い(です)」の基本形を「多い」とする。

②「多かった(です)」の基本形も「多い」とする。

③「多く(は)ない(です)」の基本形も「多い」とする。

④「多く(は)なかった(です)」の基本形も「多い」とする。

⑤：「も」「さえ」など、「は」以外のとりたて助詞が現れた場合(例：多くさえない)は、そのとりたて助詞の前で切れると考える(例：多く+さえ+ない〔形容詞+とりたて助詞+形容詞〕)。

基本形	表層語
多い	「多い(です)」「多かった(です)」「多く(は)ない(です)」「多く(は)なかった(です)」
	「多く(は)ありません(でした)」「多く(て)」「多く(は)なく(て)」「多けれ」「多かろう」「多うございます」

5. まとめ

本稿においては、助動詞「のだ」についての茶釜の解析結果に疑問を呈する

ことから始め、助動詞「のだ」、助動詞「だ」、形容動詞、動詞、形容詞について、基本形と表層語の関係を具体的に示した。仮に、日本語教育版の茶筌を作るとすると、本稿で示したような形で、基本形と表層語をリンクさせていけばいいのではないだろうか。

本稿で扱うことができた語はほんの少数であり、また、本稿で示した基本形と表層語の関係は、単なる提案であり、仮説である。したがって、今後、実際に解析を行なっていくことにより、本稿で示した提案・仮説が本当に妥当なものであるのか、検証する必要がある。

最後に、今後の研究の示唆となるであろうことを、2点述べておきたい。まず、1つめは、「日本語教師の直感に合致した文法体系を考える」ということである。茶筌で使用されている辞書は、日本語教師の直感とは合わない。「直感に合う」などと言うと、学術研究とは縁遠いもののように感じられてしまうが、しかし、それでも、「日本語教師の直感に合致した文法体系を考える」ということが重要であるように思われる。

そして、2つめは、「コンピュータによるコーパス解析を行なうことを前提・目標とした文法体系を考える」ということである。この意味するところは、従来のように、既存の文法体系からコンピュータ用の辞書を作るのではないということである。日本語をコンピュータで解析するという自然言語処理的な研究の歴史は、日本語の文法研究の歴史よりも明らかに新しい。だから、これまでは、文法研究の成果を自然言語処理的な研究に生かす形になってしまっていたが、今後は、コンピュータで解析することを前提・目標として文法体系を考えるということが、非常に重要なのではないかと思われる。

注

- 1) 「茶筌」とは、奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座(松本研究室)で開発された形態素解析システムのことである。使用希望者は、松本研究室のホームページより、無償でダウンロードすることができる。
- 2) ちなみに、「きれいだ」は、茶筌では「きれい(名詞-形容動詞語幹)+だ(助動詞)」の2語であると認識される。ここにも、茶筌と日本語教育との間の、語の単位の認定方法の違いを見ることができる。
- 3) 「のでもない」の「ない」は、助動詞であるとも考えることもあるようである。しかし、日本語教育においては、助動詞の「ない」と形容詞の「ない」を区別することは、あまり意味のないことであるように思われる。したがって、「ない」は、「助動詞」ではなく、すべて「形容詞」であると考えればいいのではないかとい

(12)

うことも提案したいと思う。

《参考文献》

スリーエーネットワーク（編）（1998）『みんなの日本語初級Ⅰ本冊』スリーエーネットワーク

スリーエーネットワーク（編）（1998）『みんなの日本語初級Ⅱ本冊』スリーエーネットワーク

追記：本研究は、文部科学省科学研究費特定領域研究「代表性を有する大規模書き言葉コーパスの構築：21世紀の日本語研究の基盤整備」（平成18～22年度）の成果の一部である。

（やまうち ひろゆき・実践女子大学教授）