

5 形態素、語、句

本節では、(直観的に)文を構成する単位と考えられる語(つまり単語)を中心に、形態素によって語がどのように構成されるか、また、語によって「句」や文がどのように構成されるかを考察する。なお、前者(つまり、形態素と語の関係)についての研究分野は形態論(morphology)、後者(つまり、語と句や文の関係)についての研究分野は統語論(syntax)と呼ばれる。

5.1 語の考え方と統計法則

語についての定義は実は難しく、いろいろな捉えかたがある。ここでは松本らの「岩波講座言語の科学3 単語と辞書」(1997)に基づいて語の「特徴」を紹介する。

- 音声的なまとまり

語は一般にひとまとまりのアクセントで発音される。「中京」も「大学」もそれぞれ語であり、標準的なアクセントではそれぞれ前半(「ちゅう」、「だい」)が低く、後半(「きょう」、「がく」)が高い。しかし、この二つが結合して「中京大学」という複合語になれば、「きょうだい」の部分が高いアクセントと持つ、というように新しい音声的なまとまりを持つようになる。

ただし、これに対する例外も存在する(「善悪」、「前学長」)

- 意味的なまとまり

語は基本的にはひとまとまりの意味概念を表す。「花見」(複合語)は、「花」と「見る」という語それぞれが持つ意味以上のものをもつ(問題:それは何だろうか、考えてみよう)。また同様に、「昔話」と「昔の話」ではかなり意味が異なる。

ただし、熟語(「カラスの行水」、「油を売る」など)のように形式的にはいくつかの語と見なされるが、意味的には全体として一つのことを指しているものもある。

- 形態的なまとまり

文を構成する要素としてみた場合に、語は形態的なまとまりを構成する。

- 句排除の制約: 語の内部に(複数の語からなる)句を埋め込むことはできない。「窓拭き」に対して「窓[一所懸命]拭き」は不適格。
- 外部からの修飾の禁止: 語の一部だけを外部から修飾することはできない。「新しい[本棚]」において、「新しい」は「本棚」を修飾し、「本」だけを修飾することはできない。
- 語彙照応の制約: 語の内部の要素を文中の照応に利用することはできない(つまり、「これ」や「それ」「あれ」で指せないということ)。「魚を釣ってそれを家に持って帰った」では「それ」は「魚」を指すが、「魚釣りしてそれを家に持って帰った」では「それ」は「魚」を指せない。

言語学でも (心理学でも) 語は母語話者の頭の中にある心的な辞書に蓄えられている項目と考えられる。

雑誌や本などの資料において出現する語を調査することがある。このような調査によって、(時代、年齢、性別などを反映して) どのような語がよく用いられるかがわかったり、語の集合により著者を特定する/特徴づける、という研究が可能になる。

そのような場合、以下の用語が用いられ、これらを知っておくことは必要。

- タイプ (type) とトークン (token)

雨は降る降る人馬はぬれる、越すに越されぬ田原坂。

という歌の一節を単語 (らしきもの) に切り分けたものが以下:

|¹ 雨 |² は |³ 降る |⁴ 降る |⁵ 人馬 |⁶ は |⁷ ぬれる |⁸ 越す |⁹ に |¹⁰
越さ |¹¹ れ |¹² ぬ |¹³ 田原坂 |

この文にはいくつ語が含まれていると考えるだろうか?

ある人は 13 語、別な人は 11 語、もう一人は 10 語 — どこが違うのか?

タイプとして考えた語を見出し語、トークンとして考えた語を単位語とも呼ぶ。

- 異なり語数と延べ語数: 資料に現れる異なる見出し語の数と単位語の数のこと。タイプとトークンを使えば、タイプ数とトークン数ともいう。
- 語彙 (lexicon): 雑誌や新聞のような言語資料に現れる見出し語の集合を、その『言語資料の上の語彙』という。
- 使用度数 (frequency): 言語資料 (H とする) における、ある見出し語 (m とする) に属する単位語の語数を、『H での m の使用度数』という。
- 使用率 (relative frequency): 言語資料 (H とする) における、ある見出し語 (m とする) の使用度数をその資料 (H) の延べ語数で割った商を、『H における m の使用率』という。
- 語彙の統計法則: 例えば使用率分布に関する Zipf(ジップ) の法則が知られている。一つの資料において、f を着目する見出し語の使用度数、r を使用度数の大きい方から振った順位とし、C を定数とすると、

$$fr \approx C \quad \dots f \text{ と } r \text{ の積はほぼ一定}$$

ちなみに Mandelbrot は以下のようなもっと近似のよい式を提案している。

$$p \approx A/(r + B)^c$$

ここで p は着目している見出し語の使用率、A、B、c はその資料に依存して決まる定数。

5.2 形態素と活用語

“unbelievable” や「ご本」のように一語でありながら、その内部をそれ自体で独立した意味を持つような単位にわけることができるものは多く存在する。また“ran” や「走ら(ない)」のように活用語では同じタイプの語でありながら異なる形態の語が存在する。

このように、語を基本単位とするのではなく、それよりも細かい単位を考えることができる。単語を構成する、一般には単語よりも小さな意味の単位を形態素(morpheme)という。

幾つかの形態素から語が構成される場合、複合、派生、豊語(「黒々」、「家々」のように々形態素の繰り返しによって作られる語)、短縮(「パーソ(ナル) コン(ピューター)」を「パソコン」とするように語の一部を切り取って作られる語)などがある。ここでは複合と派生について簡単に説明する。

- 複合(compounding): 「[腕][時計]」、「[押さえ][つける]」、「[心][強い]」のように語が複数の形態素から構成され、そのうちの 하나가全体の品詞を決定するもの。日本語の場合は後ろに来るものがその役目を持つことがほとんどである。
- 派生(derivation): 英語では先に挙げた“unbelievable”がその例である。意味の中心を担う形態素が一つあり(それを語基, base という) その前(左側)に付く形態素を接頭辞(prefix)、後ろ(右側)に付く形態素を接尾辞(suffix)という。接頭辞と接尾辞を合わせて接辞(affix)という。“unbelievable”の場合は、“un+believ(e)+able”という三つの形態素から構成され“believ(e)”が語基、“un”が接頭辞、“able”が接尾辞となる。

日本語の例としては「寒さ、強み、男っぽい、春めく、映画化、大衆的、普遍性」があげられる。英語と同様に、語基が意味の中心であるが、語全体の品詞は一番最後(右側)の要素が決定している⁹。つまり、「さ、み」は形容詞について名詞を作る接尾辞、「っぽい」は名詞について形容詞を作る接尾辞、「めく」は名詞について動詞を作る接尾辞である。

計算機で語や形態素を認識する場合、工夫しなければならないのは活用する語(品詞では動詞、形容詞、形容動詞、助動詞)の扱いである。同じ形態素であっても異なる形を取る可能性があるからである。例えば“run”の過去形は“ran”であり、“run+ed”ではない。

日本語の場合を考えると、かなり規則的な関係があることが分かる。そしてそれは、漢字仮名混じりで表記するのではなく、ローマ字(特に訓令式)で表記するとはっきりする。

以下に日本語の動詞の活用形をあげる。日本語は膠着語(agglutinative language)であり、(欧米語に見られるような時制や性などの文法的性質ではなく)後続する語によって語形が変化する可能性がある言語である。下記の活用形は、それを反映して、動詞の後ろにどのような語が来るかによって動詞の形がどのように変わるかを示したものである。(動詞の未然形1は否定の「ない」、連用形は中止用法や過去の助動詞「た」に接続する形、などを意味する)。この表に見られるように日本語の活用形は、活用形でも変化しない部分(語幹)と変化する部分(語尾)とに分けられる。

⁹ 英語の場合、“en-danger”や“a-sleep”のように接頭辞が品詞を決定するものも稀にある。

表 5.1. 日本語の動詞の活用形

動詞	未然形 1	未然形 2	連用形 1	連用形 2	終止形	連体形	仮定形	命令形
行く	ik-a (nai)	ik-o (o)	ik-i	ik- (ta)	ik-u	ik-u	ik-e	ik-e
見る	mi- (nai)	mi-(yoo)	mi-	mi- (ta)	mi-ru	mi-ru	mi-re	mi-yo, mi-ro
与える	atae- (nai)	atae-(yoo)	atae-	atae- (ta)	atae-ru	atae-ru	atae-re	atae-yo, atae-ro
来る	k-o (nai)	k-o (yoo)	k-i	k-i (ta)	k-uru	k-uru	k-ure	k-oi
する ¹⁰	s-i (nai)	s-i (yoo)	s-i	s-i (ta)	s-uru	s-uru	s-ure	s-eyo , s-iro

「行く」は国文法では五段動詞、現代の言語学では子音動詞と呼ばれる活用をする動詞の一つである。「見る」と「与える」は国文法ではそれぞれ上一段動詞、下一段動詞と呼ばれるものであるが、現代の言語学では母音動詞と呼ばれる。上の表から分かるように、かなり規則的である。

形容動詞と形容詞の活用の例も以下にあげておく。学者によっては形容動詞を認めない(名詞と考える)者もいる。ただし、名詞と決定的に違うのは、「静かだ」を例に取れば、「静かな場所」のように名詞が後接する場合に「な」という語尾をとることであり、また「に」という語尾を取って副詞的に使われること、文末に現われる場合に「だ、である、です」という語尾をとることが特徴である。なお、意味から想像できるように、命令形は存在しない。

表 5.2. 日本語の形容詞と形容動詞の活用形

	未然形	連用形 1	連用形 2	連用形 3	終止形	連体形	仮定形
寒い	samu-karo	samu-ku	samu-u	samu-kat	samu-i	samu-i	samu-kere
正しい	tadasi-karo	tadasi-ku	tadasi-u	tadasi-kat	tadasi-i	tadasi-i	tadasi-kere
静かだ	sizuka-daro	sizuka-ni	sizuka-de	sizuka-dat	sizuka-da	sizuka-na	sizuka-nara
同じ	onazi-daro	onazi-ni	onazi-de	onazi-dat	onazi-da	onazi-	onazi-nara

なお、計算機のプログラムによって、自動的に語を抽出したり、形態素に分けることを、形態素解析という。そのためのプログラムがいろいろ作られている。その中でも有名なのは、京都大学と奈良先端大学院大学で開発された JUMAN と茶筌 (Chasen) である。

課題 5 例にならって「きれい」の活用形の表を作れ。このことから、「きれい」の品詞は形容詞か、形容動詞かを判定せよ。

課題 6 以下の文章に対し、(1) 単語に分けよ、(2) それぞれの単語の品詞と(活用する語ならば)活用形を答えよ、(3) 異なり語数と延べ語数を答えよ。ただし、句読点や括弧はこの対象外である。

「古池や、蛙飛び込む水の音」は芭蕉の有名な俳句である。初案は「山吹や蛙飛んだり水の音」だったが、「飛び込む」と日常語に直したところから、わびやさびにつながる水墨画の世界が現出した。

課題 7 課題 6 の例文を茶筌 (Chasen) を用いて形態素解析せよ。茶筌の結果と課題 6 の答えとを比較し、違いについて考察せよ。

注意: Chasen は Linux でも Windows でも使えるが、パラメタ (オプション) を変えてどのような出力形式があるか、試してみてほしい。

5.3 品詞と句

以下のような観点から、語を分類するものとして、名詞 (noun)、形容詞 (adjective)、動詞 (verb)、副詞 (adverb) のような品詞が定められる。

1. 単語が意味にどのように寄与するか
2. 文の構成にどのように寄与するか

さらに、形容詞や形容動詞や動詞などの場合、上記以外に活用の規則性という点から区別されていることにも注意してほしい。

ここで「文の構成」を考えた場合、いくつかの語がまとまってある構成素 (constituent) をなす、と考えると都合がよいことが多い。その根拠としては以下のようなものがある。

1. 単一の語によって置き換え可能である (「意味が変わるかもしれないが、置き換えてもまだ文を構成する」という意味で置き換え可能)
2. 語順を変更する場合に、移動する単位となる
3. 質問やそれに対する答えの場合に省略される単位となる

いくつかの語から成り、文を構成する要素を句 (phrase) と呼ぶ。句が他の語や句とまとまってより大きな句を構成して行き、最終的に文を構成すると考える (注: 「構造」の概念)。

ここで、句 (たとえば「耳の大きな犬」) は、それより小さな句 (「耳の大きな」) や語 (「犬」) とから構成されているが、それぞれが対等の立場であるというより、どれかが (品詞の意味で) 中心的な役割を果たしており、他のものが補足的、もしくは修飾的に付いている、と考えた方がよいことが多い。そこで句の中での構成素の役割を、以下のように分類して考えよう— これは、句を構成する語の品詞によらず、どの句においても必ず主辞は一つ存在すると考える。

- 主辞 (head): それぞれの句において、それがどのように文の構成に寄与するかを決定する (品詞の意味において) 中心的な要素。句の品詞的な性質を決定するもの。

例えば「耳の大きな犬」という句は「耳の大きな」と「犬」から構成されていると考えられる。その場合、「犬」が中心的な要素であるので、これが主辞となる。また「耳の」という句は「耳」と「の」という語から構成されているが、全体として名詞を修飾する連体句として働くので、「の」が主辞と考えられる。

このように、日本語では主辞は常に句の最後の構成要素である。

- 補語 (complement): 主辞の意味を補うために必要な句の構成要素。

例えば「健が行く」という句は「健が」と「行く」から構成され、「行く」が主辞である。しかし「行く」だけでは意味的に完結しないので「健が」がそれを補っている、と考えられる。したがって「健が」は主辞「行く」に対する補語として働いている。

- 修飾語 (adjunct): 補語の意味では主辞の意味を補うものではなく、修飾的に働く句の構成要素。

例えば「ゆっくり行く」という句は、「ゆっくり」と「行く」から構成され、「行く」が主辞である。そして「ゆっくり」は、「行く」の様子を描写する修飾的な働きをしている。そこで「ゆっくり」は主辞「行く」に対して修飾語として働いていると考えられる。

(10) 句の例 (括弧内はそれぞれ句と考えられる)

名詞句	(あの (大きな 犬))
(noun phrase)	(あの (大きな ((犬 の) 首輪)))
動詞句	((犬 を) 見た)
(verb phrase)	((道 で) ((あの (大きな 犬)) を 見た))
副詞句	とても ゆっくり
(adverbial phrase)	(((((カタツムリ の) よう) に) (とても ゆっくり))

問題 8: 上の例において、それぞれの句を構成している要素 (たとえば、名詞句の例では、「あの」「大きな」「犬」) それぞれは、主辞、補語、修飾語のどの役割を果たしているか、答えよ。

5.3.1 句の構造の表現方法

例: 「公園の木」

これは「公園」「の」「木」という4つの単語が単に並んでいるのではなく、前節で見たように、「公園の」という助詞句 (役割としては連体修飾) があり、「木」と結び付いて「公園の木」という名詞句を構成すると考えられる。ここで、助詞句「公園の」においては「の」が主辞、名詞句「公園の木」において、「木」が主辞である。

この関係を以下のように括弧を用いて表すことにする。(以下では、括弧の対応関係を明らかにするため、括弧に下添字をつけて表わしている)

(11) 名詞句 (₁ 連体句 (₂ 名詞 (₃ 公園)₃, 連体助詞 (₄ の)₄)₂, 名詞 (₅ 木)₅)₁

ここで、句全体だけではなく、句を構成している要素が

『品詞 / 句の名称』 (句を構成している要素, ...)

という形で表されていることに注意しよう。

今見たように、単語が句を構成するだけではなく、句や単語からまた句が構成されるといふ階層構造¹¹ を考えることができる。この階層構造という捉え方が、文の構造ひいては文の意味を考える時に重要である。

¹¹ 「階層構造」は、自然言語理論だけではなく、いろいろな科学の分野ででてくるキーワード。

5.3.2 「曖昧さ」と句の構造を用いた曖昧さの表現

(12) 「大きな公園の木」

この文字の列には少なくとも二つの解釈が存在する。その違いは、次のように句の階層構造を示すことで明らかにできる。

(13) a. 名詞句₍₁₎ 連体句₍₂₎ 名詞句₍₃₎ 連体詞₍₄₎ 大きな₍₄₎, 名詞₍₅₎ 公園₍₅₎₍₃₎, 連体助詞₍₆₎ の₍₆₎₍₂₎, 名詞₍₇₎ 木₍₇₎₍₁₎

b. 名詞句₍₁₎ 連体詞₍₂₎ 大きな₍₂₎, 名詞句₍₃₎ 連体句₍₄₎ 名詞₍₅₎ 公園₍₅₎, 連体助詞₍₆₎ の₍₆₎₍₄₎, 名詞₍₇₎ 木₍₇₎₍₃₎₍₁₎

この例のように、複数の解釈が存在する事を、曖昧さ(あいまいさ, ambiguity)が存在する、という。

ここで取り上げた曖昧さは「構造の曖昧さ」と呼ばれるもので、文や句がどのような構成素から成り立っているか、それらがどのように結びついているかを、図示したり括弧表記することで明示することにより、その文や句がどのように曖昧かを示す事ができる。

この例の場合は、「大きな公園」が「木」と結びついているのか、「大きな」と「公園の木」が結びついているのか、ということで、二通りの曖昧さを説明できた。

「構造の曖昧さ」以外の曖昧さには、語彙の曖昧さ(例:イシ—意志、石、医師、遺志など)、文脈上の解釈の曖昧さ(例:「時計を持っていますか?」:単なる質問か、時間を知りたいのか)などがある。

5.3.3 木構造による句の構造の表現

階層構造を示す方法として、今まで見たように括弧を使う方法以外に、(グラフ理論でいう)木構造を用いる方法が良く用いられる。

(11)に対応する木構造は、図5.1のように書く事ができる。これが(11)にどのように対応しているのかを考えてみて欲しい。

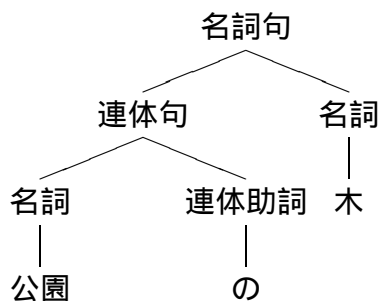


図 5.1. 「公園の木」の構造を表す木

グラフ理論における木構造についての用語の復習:

- 木とは、以下の条件を満たす有向グラフである。
 1. 「入る」枝が存在しない根 (root) と呼ばれるノードが唯一存在する。
 2. 根以外のそれぞれのノードに入る枝はただ一つである。
 3. 各ノードには、根からそれに至る道 (パス) が存在する。
- ラベル付きノード (labeled node): ここで用いる木のノードにはそれぞれラベル (上の例では、「名詞句」など) がつけられている。
- 枝 (edge) もしくはリンク (link): ノードとノードを結ぶもの。すべての枝に向きがつけられているグラフを有向グラフという。
- 根 (root node): 木にはただ一つ存在する。木を図示する場合、根を上にして書くことが多い。
- 葉 (leaf): 「出て行く」枝を持たないノード。
- 親ノード (parent node) と子ノード (child node): 枝によって結ばれた二つのノードにおいて、そのノードから枝が「出る」場合が親ノード、「入る」場合が子ノード。木を図示する場合は、上にある方が親、下の方が子であるのが普通。
- 先祖ノード (ancestor node): 任意の二つのノード v 、 w において、 v から w に至る道 (パス) が存在するならば、 v は w の先祖。 $v = w$ の場合もある。

課題 9

1. 以下のそれぞれの句に対し、句の構造を図 5.1 にならって構文木の形で表せ。(注意: これらの句の持つ曖昧さに注意せよ。何通りかに構成素をわけられる場合には、それらすべてを表し、それぞれがどのように違った意味になるかを書け。)
 - a. 黒い髪のきれいな少女
 - b. 忘れられた日記の秘密
2. 今の問題を参考にして、構文的に曖昧な文 (二通り以上の構成に分けられ、それぞれが有意味な文) を作り、その文に対して可能な構造をすべて表せ。

5.4 文節と句の違い

国文法では、文の構造を以下のようなレベルに分けて考える。

- 語—それぞれ「品詞」という同じような文法的な性質を持つ語のグループに分けられる。

国文法では品詞を二分類、すなわち、名詞や形容詞などの自立語と、助詞や助動詞などの付属語に分けて考えている。

- 文節: 一つの自立語と、それに続くいくつかの付属語(なくてもよい) からなる。

「文節」や「自立語 / 付属語」の区別は、音韻的な観点から日本語の文の構造を記述するために考えられた物であり、意味や文法的な観点から考えられたものではない。また英語など膠着語以外の言語に当てはめることが困難な概念である。

本講義では、「文節、自立語、付属語」という概念は使わず、語よりも大きな文を構成する単位として「句」を用いる。