

## 第5講 構文解析

田中重人 (東北大学文学部准教授)

### 1 パラグラフ作成課題についてコメント

#### 1.1 パラグラフの問題

- トピックセンテンスには、なるべく実質的な主張を盛り込む。「ここでは〇〇について述べる。」のような文より、「〇〇とは××である」のような文のほうがよい。
- 複数のトピックをひとつのパラグラフに混在させない。
- トピックを飛躍なく、適切な順序で並べること。内容を理解するのに必要な概念の定義や背景となる予備知識の紹介などが適切に導入されていなければならない。
- トピックセンテンスをできる限りパラグラフ冒頭に置き、それを支える文をならべてパラグラフをつくる。各文がトピックセンテンスとどのような関係にあるかを常に意識すること。

#### 1.2 文の問題

- ひとつの文にはひとつの情報だけを載せる。ふたつ以上の情報を盛り込むのは避け、できるかぎり単純な文をならべること。
- 主格・目的格・場所・時間など、必要な情報が落ちていないか確認すること。省略しなければ不自然になる場合をのぞき、なるべく省略を避ける。
- 「それ」「この」などが何を指示しているかを明確にしておくこと。

#### 1.3 ことばの表記

おなじ言葉はできる限りおなじ表記を使うこと。漢字で書くかひらがなで書くか、送り仮名のつけかたなどについて、自分なりの方針をきめて、ファイルに残していくとよい。

#### 1.4 引用・参照

文献資料等からの引用・参照に当たっては、どこからどこまでが引用・参照かをはっきりさせること。また、出典を明示することによって、読者がその文献資料を探し出すのに必要な情報をあたえなければならない。

## 1.5 初出の用語

ことばの定義をするときは、鍵括弧を使う。

……のようなものを「ら抜きことば」という。

重要なことばが外来語である場合、初出のときに原語を併記すること。

「コーパス」(corpus) とは……

## 1.6 行間

行間は、全角1文字分(下記参照)程度の高さをあけること。Microsoft Wordでは、行間を「1.5行」にすると、ほぼこの行間になる。ただし、「1ページの行数を指定時に文字を行グリッド線に合わせる」のチェックをはずす必要がある。

## 1.7 数字の表記

数量・年号・日付・時間・順序などをあらわす数字には、アラビア数字(算用数字: 1 2 3 4 …)を使う。

2人    3種類    1930年代    懲役8年

ただし、固有名詞化したものや慣用句的なもの場合には漢数字(一 二 三 四 ……)をつかってよい

十千十二支    三重県四日市市    三日月    四面楚歌

## 1.8 全角文字と半角文字

日本語の文字は、おなじ大きさの正方形の枠内にデザインされている。この正方形枠のことを「全角」という。アルファベットと数字には、全角にデザインされた文字と、半角(全角よりも幅が狭いかたち)にデザインされた文字がある。

全角文字： ひらがな    カタカナ    漢字全角

全角文字： 1 2 3 4    A B C D    a b c d

半角文字： 1234    ABCD    abcd

アルファベット／数字が2字以上つづく場合は、半角文字を使う。アルファベット／数字が1字だけの場合は、全角・半角どちらでもよい。

○ 2009年5月22日                      × 2009年5月22日

括弧類・コンマ・ピリオドなどの記号は全角・半角いずれを使ってもよい。ただし、半角記号を使う場合は、その前後のスペースを適切に入れること。

次の半角記号は、直後に半角スペースを入れる：コンマ、ピリオド、コロン、セミコロン、疑問符、感嘆符、閉じる括弧類。ただし直後にこれらの記号が続いている場合はスペースを入れない。

半角の開く括弧類は、直前に半角スペースを入れる。ただし直前が開く括弧類である場合をのぞく。

○ 大学(院)生      ○ 大学(院)生      × 大学(院)生  
○ p. 138          × p. 138

全角記号の前後にはスペースを入れない。

## 1.9 フォントの問題

文字の種類 (font) には、飾りのついたデザインのもの（たとえば明朝体）と飾りのないデザインのもの（たとえばゴシック体）がある。論文の本文には前者を用いる。後者はタイトル、セクション見出し、特に強調したい語句などにだけ使う。

Windows で使われる「MS P 明朝」は、一部の文字が全角でないサイズになっているので、使わないほうがよい。

なお、日本語文字はjis斜体*i/j* にしてはならない。また、中国語等の漢字と日本語の漢字はデザインがちがうので注意。

アルファベットにはさまざまなフォントがあるが、Roman 体かその変種が基本である。Windows では Times New Roman フォントが使える。

## 2 今回の課題

各自のパラグラフ課題の文章から、いちばん複雑な文を選んで下線を引く。その文を構文解析したうえで、わかりやすく書き直す。

## 3 構文解析とは

**文 (sentence):** 文章のなかで、句点（またはピリオド）で区切られたひとまとまりの部分。

**構文解析 (parsing):** 文の構成要素同士の修飾関係を分析すること。

**文節:** 自立語ひとつに0個以上の付属語が接続したひとまとまり。ただし、付属語とは助詞および助動詞、自立語とはそれら以外の全品詞をさす。形式体言（こと・もの……）、形式用言（ある・いる・みる……）は自立語とみなす。さ行変格活用動詞はひとつの自立語とみなす（名詞に「する」がついたものとは考えない）。複合動詞や連語はひとつの自立語とみなす。

文節に切りわけてみよう

- トンネルを抜けると、雪国だった。
- さよならだけが人生だ。

**係り受け:** 文節間の修飾—被修飾関係。

例：  
道が—歩く      月が—沈む      目で—見る  
駅に—行く      家から—出る      早く—食べる  
頭が—痛い      屋根より—高い  
大きな—手      私の—本      すごく—大きい

**構文木:** 文節間の係り受け関係を図に表したもの。修飾する（係る）文節を左に、修飾される（受ける）文節を右において、係り受け関係を線で示す

例：これらの図を使って説明すると、学生は日本文における「主語」の不在をすんなりと理解してくれる。

構文木の、いちばん右の、枝分かれしていない部分を「根」(root) という。上の例でいうと、「理解してくれる」が根である。

**並列構造:** 文のなかに、文法上同格の要素が隣り合わせに配置されていることがある。このような場合、並列の要素を上下にならべ、四角で囲んで線で区分する。

例：調査は仙台と福島でおこなった。

例：私は○○を助手席に乗せ、車を走らせた。

## 4 単純な文

構文がある程度以上複雑になると、非常に読みづらくなる：枝わかれが多い、枝が長い、並列構造のなかに複雑な枝わかれがある、枝の先に並列構造がある、係り受けや同格関係が確定できない、など。

**対策:** 余計な文節を削る、枝を切り落として独立させる、並列構造の中身を小さくする、読点などの記号を活用する、注や箇条書きや表を活用する。

# 第5講 構文解析 (追加資料)

## フォントの問題

飾りのついたデザインと飾りのないデザイン

**明朝体:** ひらがなカタカナ漢字——

**ゴシック体:** ひらがなカタカナ漢字——

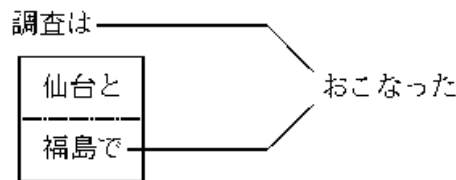
Windows で使われる「MS P 明朝」は、一部の文字が全角でないサイズになっている

**MS P明朝:** 句読点。、や「かぎ括弧」の前後が詰まる。

**MS 明朝:** 句読点。、や「かぎ括弧」の前後が詰まらない。

## 構文木の例

例：調査は仙台と福島でおこなった。



例：私は〇〇を助手席に乗せ、車を走らせた。

