

3J-7

日本語文章からの文章骨格の抽出

西村健士, 野村直之, 松本恵, 村木一至
日本電気(株) C&C情報研究所

文書作成支援環境における日本語文章からの文章骨格の抽出について述べる。まず、第1節で、本稿で対象としている文書作成支援環境について簡単に紹介し、第2節で文章骨格の役割と原文からその抽出方針について解説する。続いて、第3節で抽出処理の全体的な流れとその各ステップについて具体例を混えて説明する。

1. 文章作成からみた文章構造

ここで考える文書作成支援環境とは、文書を構成する語、文節、句、文、段落、...などの様々なレベル間にまたがって、挿入、削除、置換などの編集操作を提供するものである。また、各構成単位に対する文法的な制約を活用することによって、編集過程で生じる論理の矛盾などを検出し、文章の整合性を確認する負担からユーザを解放して、ユーザが真に創造的な作業に集中できるようにすることを目指している[野村89]。

2. 文章骨格

ユーザは、2種類の文面で文書を編集することができる。一つは原文で、もう一つは文章骨格である(図1)。文章骨格とは、原文から文章の論理構造に関連した部分を抽出したものである。文章骨格を対象とすることによって、原文章全体を表示されてもなかなか見えてこない文章の論理構造を意識的に編集することができる。原文章を編集するとその変化は文章骨格に反映されるが、逆に、文章骨格を編集するとそれに連動して原文章も変更される。

文章骨格は文章の概要把握を支援する役割もはたすが、これは要約などとは異なる。文章の要約では、受け手としてその文章を読む人が全体の要旨をつかむのに必要な情報を抽出しなければならない。一方、本稿で言う文章骨格とは、その文章を作成編集している当人が文章の流れを把握するためのものである。

骨格部分とは、基本的には、その構造単位の主要素(主題や、叙述部の中心をなす用言、その用言の必須格要素など)と、より大きなレベルの構造単位を関係づける働きをする語句(接続詞、指示詞など)である。

新しくできた研究所は、物性関連の研究の中
枢をなしており、その研究分野は機能性高分子
、電子材料、複合材料など広い範囲に及んでい
ます。また、基礎研究部門は、最新鋭の装置を
そなえ、高度な技術開発を行っておりますが、
その技術は、研究開発の成果実現に大きく寄与
しております。

図1 a (原文)

__研究所は、__中枢をなしており、その研究
分野は__範囲に及んでいます。また、基礎研究
部門は、__技術開発を行っておりますが、その
技術は、__成果実現に__寄与しております。

図1 b (骨格)

3. 骨格抽出の方法

図2に骨格抽出の手順を示す。処理は一文単位に行なう。形態素解析によって文節単位に区切られた文を単文に分け、次に、各単文を句に分ける(図3)。骨格の抽出は、単文における句の選択、単文間の関係による単文の選択、の2段階で行なう。現在のところ、これ以上の単位では骨格抽出を行っていない。以下、手順にそって、骨格抽出の方法を述べる。

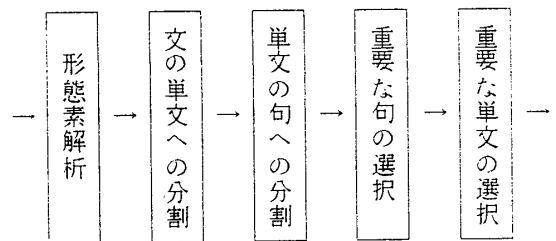


図2 処理の流れ

3. 1. 文の単文への分割

単文とは、一つの述部(用言や”名詞+「だ」”)

(文) = (単文)(単文), . . . (単文)
 (単文) = (句)(句), . . . (句)(主用言)

図3 文、単文、句の関係

を中心として、一つのまとまった叙述内容を表わすものであるとされる〔国立〕。ただし、接続助詞などの機能語を含めた述部のまとまり具合（自立性）については、その基準を設定する上で難しい問題がある。ここでは、述部が体言に係る用言の連体形である場合や独立度の比較的低い接続助詞〔亀井86〕を含む場合を除いて、用言文節を単文の述部と認めることにした。従って、単文の切れ目は以下の3つである。

- ・文の切れ目（文末）
- ・用言＋接続助詞
 ex. ～したけれども、（切れ目）～
- ・連用中止

3. 2. 各単文の主用言の選択と句への分割

単文全体をまとめる述部を主用言と呼ぶことにする。その主用言に対して、単文の残りの部分を次のどちらかに位置づける。

①単文間の関係を与える語句

②主用言に係る連用修飾句

これらを句と呼び、単文の1レベル下の単位とする。①は、接続詞とそれに相当する語句（ex. 換言すれば、早い話が、...）であって、単文やそれ以上の単位を結びつける働きをするものである。②は、通常の連用修飾句（副詞や用言の連用形）の他に、主用言の格要素となっているものも含む。

3. 3. 各単文における句の選択

各単文に対して、文章の論理構造を知る上で必要と考えられる句を選択する。

前節の①は、文章の展開を表す語句と考えられるので、すべて選択する。また、主用言と前節の②のうち、

(i)主題を表すもの

- ・ハ格や「～モ」、「～コソ」など
- ・「～とは」、「～というのは」、「～と言えば」など

(ii)主用言の必須格要素

は、文の骨格を与えるものとして選択する。その他の連用修飾句である副詞などは選択しない。

上記の方針によって骨格要素と判定された句でも、必ずしも句全体をとりあげる必要がない場合が多い。名詞は、次に挙げられるようないくつかの連体修飾句を受けている。

- ・指示詞の連体形
- ・形容詞
- ・名詞＋（格助詞、副助詞）＋「の」
- ・文（それ自身が格要素を従えている用言の連体形、または終止形）

このうち、指示詞は単文を他の単文（もしくはそれ以上の単位）と結び付ける働きを持つ語なので選択するが、その他の連体修飾句は文章の骨格に余り影響しないと判断し、原則として選択しないことにした。但し、「名詞＋の」形の連体修飾句で、実質の主名詞が”の”の前の名詞である場合は捨てることにする。

- ex. この重要で不可欠な側面をみすごして、～
 → この__側面をみすごして、～
- ex. きのお捨てた本をなくした。
 → __本をなくした。
- ex. 特色の一つとしてあげられる。
 → 特色の一つとしてあげられる。
 （「特色の」は省略されない）

3. 4. 単文の選択

最後に、単文単位で骨格として選択するか否かを判断する。例えば、同じ主題のスコープ内で並列的に連続する連用中止は、最後の用言文節のみを残しても差し支えない場合が多いと思われる。

- ex. その背景には、格闘があり、抵抗があり、
 屈服があり、敗北があった。

→

その背景には、__、__、__、敗北があった。

単文単位の骨格抽出に関しては今後も検討を続けて行きたい。

4. まとめ

文書作成環境における文章骨格の役割と、その抽出方針について述べた。今後はより上位の構造単位を対象に検討を行なう予定である。

謝辞

システムの実現の立場から示唆に富む数々のコメントを頂いた日本電気技術情報システム開発㈱の浜田氏に感謝します。また、日頃御指導いただく高島、島津の両氏に感謝します。

参考文献

- 〔亀井86〕 亀井、村木、「Lexical Discourse Grammar の提案」、信学会NLC研究会86-7
- 〔国立〕「日本語の文法（下）」、国立国語研究所
- 〔野村89〕野村他「文章執筆推敲支援システム Ide aP概要」、本大会予稿集