

接続構造解析に基づく日本語複文の分割†

武 石 英 二^{††} 林 良 彦^{††}

計算機による日本語文書の推敲支援において、長文の推敲機能を実現するためには、文を分割することによって適切な文に書き換える文分割の技術を確立する必要がある。本論文では、複文を対象とした接続構造解析に基づく文分割方式を提案する。本方式では、接続形式の導く従属句の独立性に着目することによって、述語間の接続構造を解析し、分割点となる述語を決定する。このとき、連用中止表現を形態的特徴によって分類し、接続構造の解析に利用する。これらによって、計算コストの大きい文全体の構文解析を行わずに分割すべき述語を特定することができる。また、連用中止表現を含む文に対しても、高い精度で接続構造を解析することができる。さらに、分割文を生成する際には、連用中止表現と文末述語の形態的な特徴に着目することによって、連用中止表現で接続された単文間の意味的な関係を判定し、その結果に応じた分割文を生成する。したがって、原因、方法などの関係を担う連用中止表現に対しても、適切な分割を実現することができる。最後に評価実験によって、約 86% の文に対して分割すべき述語を特定できること、約 93% の連用中止表現に対して意味的な関係を 2 個以内に限定できること、約 83% の文に対して有効な分割文を生成できることを示し、本方式の有効性を示す。

1. はじめに

近年、計算機による日本語文書の校正支援／推敲支援の研究が盛んに行われるようになってきた^{1)~3)}。その中には、書換え候補を提示する機能を目指したシステム⁴⁾もある。一方、テクニカルライティングと呼ばれる分野では、マニュアル等の技術文書を対象として、わかりやすい文章を作成するためのノウハウが検討されている^{5), 6)}。その中では、いわゆる長文は、曖昧さや紛らわしさが多くなり、読み手に負担を与えることから、最も注意すべき表現として取り上げられている。

計算機による推敲支援において、長文の書換え支援機能を実現するためには、どのような文を長文として扱うべきかを判定するだけでなく、判定された長文を分割などによって適切な文に書き換える文分割の技術を確立する必要がある。

これまでに提案されている文分割の技術は、機械翻訳のための前処理¹⁾や規格化された日本語表現の実現²⁾を目指したものが主であった。推敲支援の観点からすれば、これらの技術には、文字列レベルの情報しか用いないために適切な分割を達成する精度に問題があったり、構文解析を必要とするために計算コストが高いあるいは機械処理の早い段階で分割結果が得られ

ないという問題点があった。

本論文では、形態素解析の結果情報によって文内の接続構造がおおよそ把握できることに注目し、単文間の接続構造解析に基づく複文の分割方式を提案する。以下では、単文を接続する表現が、その表現をもつ従属句の独立性によって分類できることを示し、その分類に基づく接続構造の解析方式および分割文の生成方式を提案する。また、従来、並列的表現として扱われることが多かった連用中止表現に対しては、連用中止表現によって接続される 2 つの単文の意味的な関係を正しく同定することが、分割文の生成において重要であることを示し、その方法についても提案する。

2. 従来研究と課題

2.1 文分割の対象とする文

本論文において分割の対象とする文は、接続助詞などの表現（接続形式）によって複数の単文が接続した文（複文）である。また、これらの接続形式の前後を、独立の文として分割するような分割方法のみを考察の対象とする。連体修飾句を文として分離したり、並列句を箇条書きにしたりするような分割は扱わない。

2.2 接続構造の解析

複文を分割するには、述語間の係り受け構造（接続構造）を解析することによって、文の統語上の切れ目となる述語（分割点）を同定する必要がある。これは、具体的には、文末に係る述語のうち、もっとも文頭に近いものである。たとえば、例 1 では、接続構造が図 1 のようになっているので、分割点は「充実させてお

† Dividing Japanese Complex Sentences Based on Conjunctive Expressions Analysis by EIJI TAKEISHI and YOSHIHIKO HAYASHI (NTT Network Information Systems Laboratories).

†† NTT 情報通信網研究所

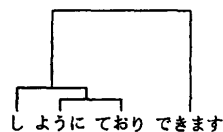


図 1 例 1 の接続構造

Fig. 1 Sentence structure of the example 1.

り」である。

例 1：本装置は、前装置をベースにし、図形などを扱えるように機能を充実させており、図形、画像データとして得られたデータも文字データと同様の扱いができます。

接続構造は、各従属句がどのような接続形式を伴っているかということと密接な関連をもっている⁷⁾⁻⁹⁾。とくに南⁹⁾は、従属句の内部における主題語、主語、修飾語などの諸要素の共存関係が文の構造的段階*を知る手がかりとなっていること、従属句は、その内部における諸要素の現れ方に基づいて 3 つの段階 (A 類、B 類、C 類) に分類できること、この従属句の分類は、たとえば B 類の従属句は、A 類、B 類の従属句を内部に含むことができるが、C 類の従属句を含むことはできないといった従属句相互の包含関係を規定することを指摘している。言い換えると、この分類は、従属句の文としての独立のしやすさを表しており、C 類、B 類、A 類の順に独立性が高いといえる¹¹⁾。

従属句の分類において、各従属句は、そこに付加された接続形式によって、「～ながら」、「～ば」のように代表されるので、従属句の分類は、接続形式の分類とみることができる (表 1)。接続形式の分類とみるのは、後で連用中止表現および接続形式「て」、「で」を他の接続形式と区別して扱うためである。また、ある接続形式で終わる従属句を、その接続形式の導く従属句

* 文におけることからの側面から陳述的側面に階層的な段階を想定することは、かなり広く受け入れられている¹⁰⁾。

表 1 接続形式分類(南⁹⁾)と助動詞分類(北原¹⁰⁾)
Table 1 Classification of conjunctive expressions and auxiliary verbs.

分類	接 続 形 式	分類	助 動 詞
A 類	ながら [継続] つつ、て、連用形	A' 類	せる、させる れる、られる
B 類	と、ても、連用形 ので、なら	B' 類	たい、ない
		C' 類	た
C 類	けれど、から、 が、て、連用形	D' 類	う、よう、だろう そうだ (伝聞)

と呼ぶ。この接続形式の分類を用いると、例 2 (a) の「ので」が B 類、「けれど」が C 類に属することから、例 2 (b) のような接続構造であることを解析することができる⁹⁾。

例 2 (a)：彼が勧めたので行ってみたけれど、それほどでもなかったよ。

(b)：[彼が勧めたので行ってみたけれど、] それほどでもなかったよ。

接続形式の分類を用いた接続構造の解析手法も、いくつか提案されている¹²⁾⁻¹⁴⁾。しかし、従来の接続形式の分類を接続構造の解析に適用する場合には、同一分類に属する接続形式が一文内に出現すると接続構造を決定することができない、連用中止表現や接続助詞「て」は複数の分類に属するため、これらを含む文では接続構造を決定することができない、形式名詞などによる接続助詞相当の語句については、「ため」、「ことによって」などを除いて明らかにされていないものが多い、などの問題点がある。

そこで、2 種類の計算機マニュアル (計算機のガイドブック、計算機システムの概説書) 中の 314 文から、25 種類、242 表現の接続形式を抽出し、接続構造の調査、単文間の入れ換えや類似した意味を表す他の接続形式への置き換えと接続構造との関連の調査、分割文の作成などを行い、マニュアル文章における複文の接続構造の特徴を分析した。3 章では、その分析結果に基づいて、次の点を特徴とする接続構造の解析方式を提案する。

- ・文構造における助動詞と接続形式の関連について考察し、従来の接続形式の分類を詳細化するとともに、その接続形式の分類をマニュアル文章に出現する接続助詞相当の語句へも拡張する。
- ・従来の接続形式の分類では複数の分類に属する連用形による接続や接続助詞「て」による接続を、それらの形態的な特徴に着目して分類し、その分類に基づく接続判定ルールを用いて接続構造を解析する。

2.3 分割文の生成

適切な分割文を生成するためには、分割が可能であるか否かを判定するだけでなく、分割が可能である場合には、分割後の文の順序、分割点の述語表現の変換、主題表現などの文要素の補完や移動、分割後の文間に補う接続詞表現などを決定する必要がある (これらをまとめて分割パターンという)。このとき、連用中止表現によって接続された文間には、原因/理由、

手段／方法といった接続関係（連用中止表現の意味用法）が見られることが指摘されていること^{7), 8), 15)}、マニュアル文章を対象にした場合にも、このような接続関係をもつ連用中止表現が比較的多く出現することから、連用中止表現を分割点とする文の分割パターンを決定するためには、意味用法を特定することが必要であると考えられる。

たとえば、例3のような文分割では、分割点の述語表現に「ている」を補う操作、分割後の文間に接続詞表現「それによって」を補う操作が必要である。そのためには、「把握分析し」を並列接続として扱うのではなく、因果的な接続関係をもつことを認定しなければならない。

例3: Aシステムでは、時代の動向を的確に把握分析し、ハードウェアとソフトウェアの計画的な開発を進めています。

→Aシステムでは、時代の動向を的確に把握分析しています。

それによって、ハードウェアとソフトウェアの計画的な開発を進めています。

従来、提案されている文分割方式は、分割点以前を第1文、分割点以降を第2文として分割する、接続形式ごとに分割時に補う接続詞表現をあらかじめ1つに決定しておく、そのとき連用中止表現や接続助詞「て」による接続は並列接続として扱う、という方式が多いことから^{1), 2)}、適切な分割を達成する精度に問題があると考えられる。

本論文では、4章において、次の点を特徴とする分割文の生成方式を提案する。

- ・分割点の接続形式、文頭の接続詞表現や主題表現などの分割点前後の文節パターンから分割パターンを決定する。
- ・連用中止表現とその係り先の述語表現の形態的特徴に着目することによって連用中止表現の意味用法を判定し、その意味用法に基づいて分割パターンを決定する。

3. 表現形態分類に基づく複文の接続構造解析

3.1 接続の機能をもつ語句の分類

接続形式を図2のように分類する。接続形式のうち連用形接続（助動詞などの連用形を含む）および接続助詞「て」、「で」による接続をあわせて連用中止表現

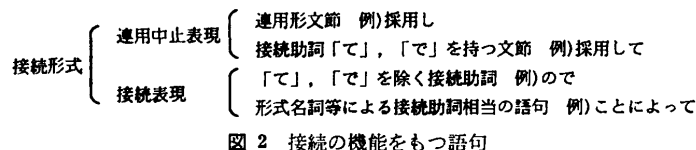


図2 接続の機能をもつ語句

Fig. 2 Types of conjunctive expressions.

と呼ぶ。連用中止表現以外の接続形式、すなわち「の」「から」などの接続助詞および「ため」、「ことによって」などの接続助詞相当の語句をあわせて接続表現と呼ぶ。また、述語表現とは、いわゆる述語と接続形式をあわせた表現を指すものとする。

たとえば、2章の例1は5つの述語表現をもつ。そのうち「し」、「扱えるように」、「充実させており」が接続形式をもち、「し」、「充実させており」が連用中止表現、「ように」が接続表現である。また、「得られた」が連体修飾の述語、「できます」が文末の述語である。

3.2 接続表現の分類と接続構造解析への適用

北原¹⁶⁾は、承接関係によって助動詞を分類し（表1）、文構造との関係について考察している。そこでは、承接関係による助動詞の分類が文構造の段階とよく対応していること、とくに接続形式分類をA類～C類、助動詞分類をA'類～D'類とすると、接続形式の導く従属句にどの助動詞が現れうるかによって、表1のようにA類とA'類、B類とB'類、C'類、C類とD'類が対応することを指摘している。

接続形式分類と助動詞分類は、B類がB'類、C'類の2分類と対応すること以外の部分では、文構造を説明する理論としてよく一致していること、助動詞分類B'類とC'類には、文構造における役割に違いが見られること*から、接続形式B類を、B'類に対応するものとC'類に対応するものの2つに分類し、接続構造の解析に利用することが考えられる。

一方、形式名詞などによる接続助詞相当の語句も、構文的には接続助詞と同等の機能をもち、接続助詞と同様に語句の種類によって従属句としての独立性に違いが見られる。したがって、接続助詞相当の語句も接続助詞と同一の扱いができることが望ましい**。しかし、どのような表現を接続助詞相当の語句として扱うべきかは、対象とする文章の種類によって異なることから、マニュアル文章を対象としたデー

*「ずいぶん」、「もっと」などの修飾する内容が程度的概念である修飾成分が修飾することができるのは、希望や打ち消しなどB'類の助動詞で表される付属概念までであって、C'類の助動詞で表される過去概念は、その修飾対象の範囲に入らない¹⁶⁾。

**一部の接続助詞相当の語句が接続助詞に準じて扱えることは、南¹⁷⁾も指摘している。

表 2 接続表現分類

Table 2 Proposed classification of conjunctive expressions.

分類	接 続 表 現 例
I 類	ながら[継続], つつ, まま
II 類	ことにより/ことによって[方法], ば よう(に), とき(に), まで(に), ても
III 類	のに, ので, なら, とともに, がゆえに ため(に)[原因], にもかかわらず
IV 類	し, けれど(も), から, が

タの分析結果に基づいて, 接続助詞相当の語句として扱うべき表現を抽出するとともに, それらがどの接続形式分類に相当するかを整理した。

上述した接続形式分類の細分類および接続形式分類の接続助詞相当語句への拡張に基づいて, 接続表現(文献 17) およびデータ分析結果をもとに抽出した 85 表現)を 4 つに分類した。その一部を表 2 に示す。この接続表現分類を用いることによって, たとえば III 類の接続表現によって導かれる従属句は, IV 類の接続表現によって導かれる従属句の一部になるが, I 類, II 類の接続表現によって導かれる従属句の一部にならないという従属句間の包含関係を決定する。また, 接続表現によって導かれる従属句が連体修飾句の一部になりうるかも, この接続表現分類を用いて判定する。すなわち, I 類, II 類の接続表現によって導かれる従属句は連体修飾句の一部となりうるが, III 類, IV 類の接続表現によって導かれる従属句は連体修飾句の一部にならない⁹⁾。

3.3 連用中止表現の分類と接続構造の解析

従来の接続形式分類では, 連用中止表現は, 複数の分類に属する表現として扱われてきた。これは, 連用中止表現が, その接続先の述語表現を構成する助動詞や補助用言などの付属表現を共有することによって, 付属表現を省略した形態で使用されるため, 文構造の段階を一意に決定することができないことに

表 3 連用中止表現の分類例

Table 3 Classification of "Renyoun-tyuushi" expressions.

分類	表現形態分類	表 現 例
a 類	し型 させ型 して型	動作動詞の連用形 「せる/させる」の連用形 a 類+「て」
b 類	して型 あり型 く型	動作動詞の連用形+ていて 状態動詞の連用形 形容詞の連用形
c 類	で型	「だ」の連用形
d 類	であり型 ており型	「である」の連用形 「ておる」の連用形

る^{8), 9)}。

しかし, 連用中止表現を 1 つの接続形式として扱うのではなく, 付属表現の有無などの形態的な特徴に着目して連用中止表現を分類すれば, その分類を接続構造の解析に利用することができると考えられる。実際にマニュアル文章では, 動詞の連用形に限らず, 形容詞, 助動詞, 補助用言から構成される連用中止表現も多い。さらに, 「プロセッサ同士を結合し, ファイルの共有ができる本体装置間結合装置があります。」のように, 「～し」は連体修飾句の一部となるが, 「～ており」は連体修飾句の一部とならないことから, 連用中止表現の形態と接続構造には密接な関連があると考えられる。そこで, 連用中止表現の形態の特徴および連用中止表現の導く従属句の独立性から, 連用中止表現を表 3 のように分類した。

マニュアル文章の分析によると, 接続形式から連用中止表現への接続の可否および連用中止表現から接続形式への接続の可否は, この分類および接続表現分類の 2 項関係として決定できることが多い。データ分析の結果に基づいて作成した接続判定ルールを図 3 に示す。図 3 において, たとえば「(a) 接続表現→連用中止」は, 接続表現の連用中止表現への接続を判定するルールである。「II 類→d 類」が「○」であるとは, II 類の接続表現が d 類の連用中止表現を越えて後部の述

	a 類	b 類	c 類	d 類
I 類	?	?	?	○
II 類	?	?	?	○
III 類	?	?	?	○
IV 類	x	x	x	?

(a) 接続表現→連用中止
(a) Conjunctive expression
→"Renyoun-tyuushi".

	I 類	II 類	III 類	IV 類
a 類	?	?	?	○
b 類	x	?	?	○
c 類	x	x	x	○
d 類	x	x	x	x

(b) 連用中止→接続表現
(b) "Renyoun-tyuushi" →
conjunctive expression.

	a 類	b 類	c 類	d 類
a 類	?	?	○	○
b 類	x	?	○	○
c 類	x	x	?	○
d 類	x	x	x	?

(c) 連用中止→連用中止
(c) "Renyoun-tyuushi" →
"Renyoun-tyuushi".

注) ○: 接続可, x: 接続不可, ?: 判定不可能。

図 3 連用中止表現の接続判定ルール

Fig. 3 Analysis rules for "Renyoun-tyuushi" expressions.

語表現に接続することはない、すなわちⅡ類の接続表現は、d類の連用中止表現もしくはそれよりも前方の述語表現に接続することを表している。「Ⅳ類→a類」が「×」であるとは、Ⅳ類の接続表現によって導かれる従属句は、a類の連用中止表現によって導かれる従属句の一部にならない、すなわちa類の連用中止表現よりも前方もしくは後方の述語表現に接続することを表している。

また、連用中止分類は、従属句としての独立性を反映したものであり、a類、b類、c類の連用中止表現によって導かれる従属句は、連体修飾句の一部になりうるが、d類の連用中止表現によって導かれる従属句は、連体修飾句の一部にならない。

3.4 接続構造候補の絞り込み

述語表現間の距離、読点の有無、述語の品詞や動詞性などの文法的属性も、述語表現間の係りやすさ、係りにくさを規定する要因となっている^{12),13)}。これらを、係り元の述語表現、係り先候補の述語表現、両者の位置的な関係などを適用条件とするルールとして整理した。その一部を表4に示す。表4において、「係りやすさ」とは、適用条件を満たす述語表現間の係り受けが成立しやすいか(+), 成立しにくい(-)を示している。また、「優先度」とは、ルールによって係りやすさや係りにくさを規定する強さに違いが見られることから、係り受けを規定する強さによって各ルールを3つに分類したものである。たとえば、表4では、「ことによって」を含む述語表現が形容詞や形容動詞に係りにくいというルールが、読点の有無によるルールよりも強く働くことを示している。

接続表現分類および連用中止表現によるルールを適用しても接続構造を一意に決定することができない場合には、接続構造候補ごとに適用条件を満たすすべてのルールを抽出し、ルールの個数、それぞれのルールの「係りやすさ」や「優先度」によって接続構造候補

計算機を導入し、これらの計算機をネットワークで接続し、高度化する傾向があります。

↓
1位: A (j1→j3, j2→j3) [②+⑤+⑤]
2位: B (j1→j4, j2→j5) [②+⑤]
3位: C (j1→j2, j2→j3) [②+③+④+⑤+⑤]
4位: D (j1→j4, j2→j4)
5位: E (j1→j2, j2→j4) [③+④]

図4 接続構造候補の順位付けの例

Fig. 4 An example of sentence structure disambiguation.

間の順位付けを行う。以下では、図4を例として候補間の順位付けの方法を述べる。図4において、[]内は表4のルールのうち適用条件を満たすものの番号を表している。

まず、適用条件を満たすルールのうち、優先度「大」のルールの個数を比較し、個数の多い候補を優先する。ただし、ルールの「係りやすさ」が「-」の場合は、-1個として算入する。これは、後述する優先度が「中」、「小」のルールについても同様である。図4では、優先度「大」のルールの適用条件を満たす候補がないため、順位付けを行うことはできない。

次に、優先度「中」のルールの個数を比較し、個数の多い接続構造を優先する。図4では、3つの候補A, B, Cがルール②の適用条件を満たし、2つの候補C, Eがルール③の適用条件を満たす。ルール③の「係りやすさ」は「-」なので、適用条件を満たすルールの個数は、候補Aと候補Bが1個、候補Cと候補Dが0個、候補Eが-1個となり、1位から2位、3位から4位、5位の3つのグループに分けることができる。

最後に、優先度「小」のルールの個数を比較する。図4では、候補Aの2組の係り受け関係がルール⑤の適用条件を満たし、候補Bの1組の係り受け関係がルール⑤の適用条件を満たす。ルール⑤の「係りやすさ」が「+」なので、候補Aが1位、候補Bが2位となる。また、候補Cでは、1組の係り受け関係がルール④の適用条件を満たし、2組の係り受け関係がルール⑤の

表4 述語間の係りやすさを判定するルール例
Table 4 Examples of disambiguation rules.

No.	適用条件			係りやすさ	優先度
	係り元述語表現	係り先述語表現	述語間の関係など		
①	ことによって[方法]	形容詞, 形容動詞	—	—	大
②	連用中止 a類	動作性動詞	文節が隣接	+	中
③	—	連用中止 a類	係り先述語が条件②を満たす	—	中
④	読点を含む	読点を含まない	述語が隣接	+	小
⑤	連用中止	係り元述語と同一種類	—	+	小

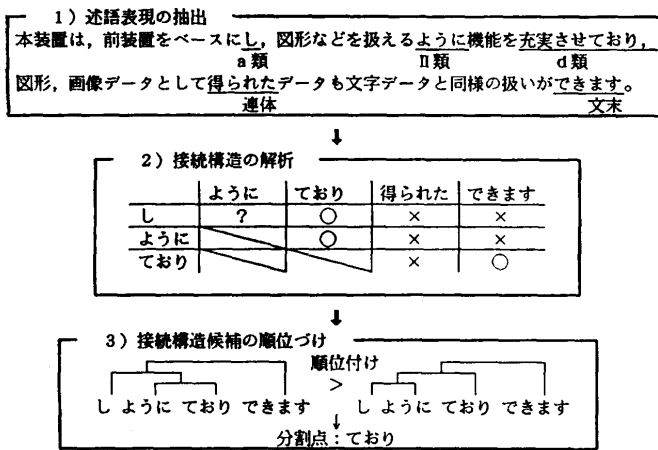


図 5 例 1 に対する接続構造解析例

Fig. 5 Conjunctive expressions analysis for the example 1.

適用条件を満たす。ルール④、⑤の「係りやすさ」は「+」なので、候補Cが3位、候補Dが4位となる。

3.5 解析例

2章の例1に対する接続構造の解析過程を図5に示す。「し」がa類、「ように」がII類、「充実させており」がd類、「得られた」が連体修飾の述語、「できます」が文末の述語である。d類の連用中止表現が連体修飾句の一部にならないという制約および図3の接続判定ルールを用いると、述語表現間の係り受け関係は図5, 2) のようになる。たとえば、1行目については、図3の接続表(b)を用いると「し→ように」は「?」であり、また図3の接続表(c)を用いると、「し→充実させており」は「○」である。したがって、「し」は「ように」、「充実させており」のいずれかに接続し、「得られた」にも「できます」にも接続しない。

図5, 2) からは、2通りの接続構造が考えられる。これらに対して、表4のルール④を用いると、図5, 3) のように順位付けすることができる。ただし、分割点は、いずれの場合も「充実させており」になる。

4. 分割文の生成

4.1 分割点が接続表現の場合の分割パターン

従来から指摘されているように、複文を分割することができるかどうか（分割

の可否）やどのように分割したらいいか（分割パターン）は、主として分割点の接続形式によって決定できることが多い^{11), 18)}。さらに、推敲支援を目的とした場合には、原文の内容を損なわずに分割するだけでなく、わかりやすさなどのマニュアル文章としての適切性についても考慮する必要がある。マニュアル文章における文分割の特徴を分析した結果を表5に示す。ただし、分割の対象としている複文とその前後の文との結びつきや主題の連鎖などの文脈的要因については、考察の対象外とした。表5の「分割可能」における「順序不変」とは、分割点以前の部分と分割点以降の部分の順序を入れ換えずに分割できる

ものであり、「順序交換」とは、順序の入れ換えが必要なものである。また、「両者可能」とは、両者の分割パターンによって分割することができるものである。データ分析の結果から、次のことがわかった。

1) 分割の可否、分割パターンは、分割点の接続形式によってほぼ決定することができる。表5において、目的を表す「～ため(に)」を分割点とする文のうちの「分割不可」の4文は、「この」や「次の」のような

表 5 マニュアル文章における複文の分割パターン
Table 5 Investigation of sentence division patterns in manual texts.

接続属性	分割点の接続表現	分 割 可 能				分割不可
		順序不変	順序交換	両者可能	「A+[接続表現], B.」の分割パターン	
逆接	が	4	0	0	A. しかし, B.	0
原因	ので	7	0	13	A. したがって, B. B. これは, Aからです.	0
	ため(に)	0	0	10	A. そのため, B. B. これは, Aからです.	0
並列	とともに	3	0	0	A. それとともに, B.	0
	ほか	1	0	0	A. また, B.	0
	たり	0	0	0	—	4
目的	ため(に)	0	6	0	B. これは, Aためです.	4
	よう(に)	0	0	0	—	3
条件	ば, と, たら	0	0	0	—	21
逆条	ても	0	0	0	—	3
とき	とき(に), まえ(に), 等	0	0	0	—	17
方法	ことで, ことによって	0	0	0	—	8

「小型機は、端末と同程度の温室度条件で動作できるように設計しているので、分散プロセッサや小型ホストとして、一般オフィスに設置できます。」

<分割ルール>

入力パターン	分割パターン	形態素操作
Nは V ₁ ので V ₂ .	(P1) Nは V ₁ . したがって、 V ₂ .	－ので
	(P2) Nは V ₂ . これは、 V ₁ からです.	－ので

↓

- (P1)小型機は、端末と同程度の温室度条件で動作できるように設計しています。したがって、分散プロセッサや小型ホストとして、一般オフィスに設置できます。
- (P2)小型機は、分散プロセッサや小型ホストとして、一般オフィスに設置できます。これは、端末と同程度の温室度条件で動作できるように設計しているからです。

図 6 分割ルールとその適用例

Fig. 6 An example of sentence division rule and its application.

指示表現を含む文であり、このような表現の多くは、形態素解析の結果情報によって抽出することができる。また、「～ので」を分割点とする文のうち、「順序不変」の7文中5文は、「～ので、～ください。」のパターンである。これも形態素解析の結果情報によって抽出することができる。

2) 文要素の移動が必要なものは、「ので」や「ため」を含む文の分割のように、分割点前後の順序の入れ換えを必要とする場合が多い。表5では、そのような文要素は文頭の接続詞(8文)と主題表現「～ハ」(12文)だけである。

3) マニュアル文章に出現する接続詞の種類はかなり限られており(13種類, 49表現)、同一の意味に対する接続詞が複数ある場合にも、実際に使用される接続詞は少数に限定することができる。

これらのことから、分割点の接続表現、接続詞の有無や主題表現の有無などの分割点の前後の文節パターン(これらをまとめて入力パターンという)によって、分割の可否や分割パターンを決定できる場合が多いと考えられる*。そこで、図6のように、入力パターンとその分割パターンの対を分割ルールとして記述しておき、入力パターンにマッチする分割ルールの分割パターンに従って分割候補文を生成する。

4.2 連用中止表現の意味用法判定と分割パターン

(1) 意味用法の判定

連用中止表現の意味用法については、これまでいくつかの分析がなされている^{7),15)}。しかし、それらは、連用中止表現とその係り先の述語との並行性や連用形接続と「～て」との違いについて考察したものが多

い。そのため、連用中止表現の意味用法の判定にそのまま適用することはできない。

マニュアル文章から抽出した連用中止表現(109表現)の意味用法を分析した結果、次のような特徴をもつことがわかった。

1) 可能を表す表現を含む述語に係る連用中止表現は、原因の意味を担うことが比較的多い(37表現中11表現)。これは、可能を表す表現がその前件に原因や理由となる記述を要求することが多いことによると考えられる。

2) 「～で」や「～であり」の形の連用中止表現は、因果的な関係(原因や根拠)を担うことが少なく、添加や対比の意味用法であることが多い(15表現中12表現)。

3) 「～し～する」や「～ており～している」など、連用中止表現がアスペクトなどが同一な述語に係る場合には、因果的な関係を担うことは少なく、対比や添加の意味用法であることが多い(39表現中31表現)。これは、このような係り受け関係が、連用中止表現の基本的な文型である^{8),9)}からだと考えられる。

これらのことから、連用中止表現とその係り先の述語表現の形態的な特徴に着目することによって意味用法を判定する。接続表現への書換えおよび文分割の観点から意味用法を表6のように分類した。表6において、「-」は文分割が不適切であることを示している。また、表3の連用中止分類との形態的な並行性という観点から、係り先の述語表現を表7のように分類した

表 6 意味用法の分類
Table 6 Semantic usages of "Renyoun-tyuushi" expressions.

用法	接続表現への変換	分割時の接続詞
対 比	～, また	また
添 加	～, そして	そして
原 因	～ので,	したがって/そのため
方 法	～ことによって,	—
根 拠	～ことによって,	それによって

表 7 述語表現の分類
Table 7 Classification of predicates.

ている型	動作動詞+ている
できる型	動作動詞+できる
する型	その他の動作動詞
ある型	状態動詞
形容詞型	形容詞
ダ文型	名詞述語文

* 入力パターンから分割パターンを一意に決定できない場合には、複数の分割パターンを用意しておき、複数の分割候補を出力することとした。分割パターンを一意に決定することができない原因は、文脈的な要因の他、接続詞の選択に関する語用論的な制約、接続詞や副詞のスコープの問題などであり、今後の課題である。

表 8 連用中止表現の意味用法ルールの例
Table 8 Examples of semantic usage rules for
“Renyō-tyuushi” expressions.

	する型	ている型	できる型
し型	①添加 ②原因	①添加 ②根拠	①添加 ②原因
して型	①添加 ②方法	①方法 ②添加	①方法 ②添加
あり型	①添加 ②原因	①添加 ②根拠	①原因 ②添加
でき型	①原因 ②添加	①根拠 ②添加	①対比 ②原因
であり型	?	①添加	①添加 ②原因
ており型	?	①根拠 ②添加	①原因

(以下、述語分類という)。これらの分類に基づいて、マニュアル文章の分析結果から作成した意味用法判定ルールの一部を表 8 に示す。これらのルールは、意味用法を一意に決定するのではなく、2 個程度の意味用法に限定するものであり、「?」は連用中止表現による接続として不適切だと考えられるもの⁸⁾である。

(2) 分割パターン

マニュアル文章から収集した連用中止表現の分析によると、分割点が連用中止表現である場合には、分割点の前後の順序の入れ換えや主題表現などの移動を必要としない場合が多い。また、連用中止表現の文末の述語への書換えでは、意味用法に応じた書換えを必要とする場合がある。たとえば、分割点「し型」の連用中止表現で、文末の述語が可能を表す表現を含む場合には、意味用法に応じて 2 通りの書換えを必要とする場合がある。すなわち、意味用法が「原因」のときには、「ている」を補ったほうが適切である場合が多く、意味用法が「添加」のときには、文末の述語と同様に可能を表す表現を補ったほうが適切である場合が多い。

これらのことから、分割点が連用中止表現の場合の分割パターンを、連用中止分類、意味用法およびその前後の文節パターンによって決定する。分割ルールの例を表 9 に示す。たとえば、2 章の例 1 では、3 章で示したように分割点が「充実させており」(ており型)であり、文末の述語表現が「できます」(できる型)であることから、表 8 によって意味用法は「原因」と判定される。さらに、表 9 によって、接続詞「したがって」が選択され、例 4 のような分割文を生成する。

例 4: 本装置は、前装置をベースにし、図形などを扱えるように機能を充実させています。
したがって、図形、画像データとして得られたデータも文字データと同様の扱いができます。

表 9 連用中止表現に対する分割ルールの例
Table 9 Examples of sentence division rules
for “Renyō-tyuushi” expressions.

分割点	文末	用法	接続詞	形態素操作
し型	できる型	添加	そして	+できる
		原因	そのため	+ている
ており型	—	原因	したがって	-ており +ている
であり	—	添加	—	-である +です

表 10 接続構造解析の評価結果
Table 10 Results of the conjunctive expressions
analysis.

正 解	候 補 数			
	2	3	4 以上	計
一 意	12/ 6	3/ 2	4/ 1	19/ 9
1 位	36/24	10/ 9	8/ 7	54/40
2 位	5/ 4	1/ 2	2/ 1	8/ 7
3 位	—	0/ 0	0/ 1	0/ 1
4 位	—	—	0/ 2	0/ 2
計	53/34	14/13	14/12	81/59

注) ルール作成に使用した文/それ以外 (ともに文の数)。

表 11 分割点判定の評価結果
Table 11 Results of the identification of sentence
division points.

正 解	候 補 数			
	2	3	4	計
一 意	16/ 8	4/ 1	3/ 1	23/10
1 位	45/32	3/ 7	0/ 1	48/40
2 位	7/ 6	3/ 1	0/ 0	10/ 7
3 位	—	0/ 2	0/ 0	0/ 2
計	68/46	10/11	3/ 2	81/59

注) ルール作成に使用した文/それ以外 (ともに文の数)。

5. 評 価

5.1 評価結果

本論文で提案した接続構造の解析方式、分割文の生成方式の有効性を確認するために、2 種類の計算機マニュアルを対象として評価実験を行った。

接続構造解析の評価は、2 通り以上の接続構造が考えられる文のうち、ルール作成に使用した文から 81 文、それ以外の文から 59 文を対象とした、評価結果を表 10、表 11 に示す。表 10 における「候補数」とは、非交差条件を満たす接続構造の総数であり、表 11 における候補数とは、分割点候補の総数 (連体修飾句を

表 12 接続構造解析の正解率と多義解消率
Table 12 Accuracy of the conjunctive expressions analysis.

対 象	候 補 数			
	2	3	4以上	平 均
ルール作成に使用した文 上記以外	90.6/90.6 88.2/88.2	92.9/96.4 84.6/92.3	85.7/91.1 66.7/85.7	90.1/91.7 84.5/88.6

注) 正解率/多義解消率.

表 13 分割点判定の正解率と多義解消率
Table 13 Accuracy of the identification of sentence division points.

対 象	候 補 数			
	2	3	4以上	平 均
ルール作成に使用した文 上記以外	89.7/89.7 87.0/87.0	70.0/85.0 72.7/77.3	100/100 100/100	87.7/89.5 84.7/85.6

注) 正解率/多義解消率.

表 14 連用中止表現の意味用法判定の評価結果
Table 14 Results of the semantic usage identification for "Renyō-tyuushi" expressions.

正 解	対 象 表 現	
	ルール作成に使用 した表現	左記以外の表現
1位が正解	67 (77.9%)	57 (56.4%)
2位が正解	15 (17.4%)	30 (29.7%)
3位が正解	2 (2.3%)	1 (0.9%)
正解なし	2 (2.3%)	13 (12.9%)
計	86	101

含む場合には候補なしも総数に含む)である。正解欄の「一意」とは、接続構造、分割点が一意に決定したものである。その他の正解欄は、複数の接続構造、分割点が出力された文に対して、正解が何位であることを示している。

表 10, 表 11 から求めた正解率, 多義解消率を表 12, 表 13 に示す。正解率は一意または1位の文の数の割合である。多義解消率は、候補数 m , 正解 n 位の文(一意に決定した文は、1位とした)の多義解消率を

$(m-n/m-1) \times 100(\%)$ として、候補数ごとに平均をとったもので、多義が解消された割合を示す目安である。

連用中止表現の用法判定の評価は、ルールの作成に使用したものから 86 表現, それ以外のものから 101 表現を対象とした。そこには、分割点であるもの、分割点でないものの両者を含んでいる。評価結果を表 14 に示す。表 14 における「正解」とは、複数の意味用法が出力された場合(最大3個)の正しい意味用法の順位を示している。

分割文の評価は、ルール作成に使用した文から 104 文, それ以外の文から 112 文を対象とした。接続構造が一意

に決定しない場合には、3 位までの接続構造候補に対して分割文の生成を行った。また、複数の分割パターンがありうる場合や分割点の連用中止表現の意味用法が一意に決定しない場合には、複数の分割文を出力するものとした。評価結果を表 15 に示す。

表 15 において、「有効」とは、前後関係などを考慮せずに1文の範囲で見たときに、分割文が原文の意味をほぼ反映していると判断した文である。「失敗」とは、出力された分割文が不適切だと判断した文であり、その原因別に3つに分類した。そのうち、「接続構造」とは、正しい接続構造が4位以下であるため、適切な分割点が出力されなかった文、「意味用法」とは、意味用法の判定において正しい意味用法が出力されなかった文、「分割ルール」とは、決定された分割パターンが不適切な文である。ここで、分割点が接続表現の場合の意味用法欄は、「ため」(目的, 原因)や「が」(逆接, 前置きなど)のように複数の意味を表す接続表現*

* これらの接続表現に対しては、接続構造を解析する際に、ヒューリスティック¹¹⁾を用いて多義を解消している。

表 15 分割文生成の評価結果
Table 15 Results of the divided sentence generation.

分 割 点	分 割 結 果						
	有 効	失敗（原因別）			小 計	非対象文	計
		分割ルール	意味用法	接続構造			
接続表現	32／28	1／ 5	6／ 1	0／ 0	39／34	0／ 0	39／ 34
連用中止表現	49／42	2／ 5	5／ 5	0／ 1	56／52	9／26	65／ 78
計	81／70	3／10	11／ 6	0／ 1	95／86	9／26	104／112

注) ルール作成に使用した文/それ以外 (ともに文の数)。

に対する値である。また、非対象文とは、接続構造が一意に決定されなかったため、分割点が存在しないにもかかわらず分割文が生成された文のことである。したがって、表 15 における合計数が、分割文が生成された文の総数であり、そこから非対象文を除いたもの(小計)が分割文生成の評価対象とすべき文の数である。

5.2 考 察

(1) 接続構造解析および分割点判定

表 10, 11 からわかるように、接続形式分類による接続構造の解析能力は、接続構造を一意に決定するという観点からはあまり高いとはいえない。接続構造を一意に決定できないのは、連用中止 a 類および連体修飾句によるところが大きい。しかし、正解が 4 位以下である文がないこと、正解率がルール作成に使用した文で約 90%、それ以外の文で約 83% と高い値であること、多義解消率が 90% 前後の値であることから、3 章で提案した接続構造の解析方式の有効性を確認することができる。とくに、a 類以外の連用中止表現を含む文の解析精度が高いことや、表 12 のルール作成文以外において正解率と多義解消率の差が大きいことは、接続構造の解析効果によるものといえる。

順位付けによる接続構造の多義の解消効果については、正しい接続構造や正しい分割点の多くを 1 位に順位付けていることから、一定の有効性をもつことがわかる。しかし、ルール作成文とその他の文での評価結果に差があることから、ルールの数や適用条件の整備が不十分であることがわかる。さらに、ルールの充実、適用条件の詳細化を図るとともに、順位付けのアルゴリズムについてもさらに検討する必要がある。

3 章で提案した接続構造の解析方式を拡張し、解析精度を向上させる手法として、述語の意味的な情報やマニュアルにおける主題表現、指示表現、述語表現などの使用形態に関する情報を導入することが考えられる。その際、文章構造の解析手法として提案されている述語の連鎖¹⁹⁾や単文によって含意される命題の連鎖^{20), 21)}に着目する手法が参考になると考えられる。とくに、これらは、連用中止 a 類のような形態的な特徴に乏しい表現や連体修飾句を含む文の解析、接続構造の多義解消のためのルールの詳細化に有効であると考えられる。

(2) 連用中止表現の意味用法判定

表 14 において、1 位または 2 位が正解のものを合わせると、ルール作成表現で約 95%、それ以外で約

86% であることから、連用中止表現の意味用法は、かなりの精度で 2 個程度に限定できること、その際に 4 章で提案した意味用法の判定方法が有効であることがわかる。しかし、「～し」のように形態的な特徴が少ない表現の意味用法を一意に特定するためには、(1) で考察したように、述語の意味的な情報を導入したり、可能や禁止などを表す表現と意味用法との関係について分析し、判定ルールを拡張、詳細化することが必要だと考えられる。

また、ルール作成表現とそれ以外の「正解なし」の件数の差が大きいにみえるが、ルール作成に使用した表現以外における「正解なし」の 13 箇所のうち 12 箇所は、構文的には副詞的な役割を果たしているものである。これらの表現は、「効率よく」(5 箇所)、「一貫して」(2 箇所)など固定的なものが多いため、助詞相当の表現あるいは副詞相当の表現として辞書登録することによって対処することができると考えられる。

(3) 分割文の生成

「～し」などの連用中止表現に対する意味用法判定の失敗や分割パターンの不備が、分割文生成の失敗の大部分を占めている。したがって、(1), (2) で考察した接続構造解析、意味用法判定の精度の向上は、分割文生成の精度の向上にとっても重要であることがわかる。

一方、文要素の移動や補完の不備による誤りは少ない。文の構成要素を移動する必要があるのは、分割点よりも前の要素が文末の述語に係る場合である。しかし、分割が可能な接続形式は独立性の高い表現が多いため¹⁰⁾、主題表現や文副詞など以外の要素に対しては、不自然な文になったり、読み手に負担を与えたりする。評価対象文でも、そのような要素はほとんど存在しなかった。また、文要素を補完する必要があるのは、格要素の省略(ゼロ代名詞化)がある場合である。ゼロ代名詞化は主題表現や主語が大部分であり、そのほとんどが同一文内または直前の文に先行詞が現れることが指摘されている²²⁾。評価対象文でもほとんどが主題表現であること、分割の際に補完が行われなくても補完すべき要素は直前の文にあることから、不自然な文になることは少なかった。

また、分割後のそれぞれの文の長さや両者のバランスを考慮していないために、不自然な文になることがある。文の長さや文のもつ情報量の観点などから分割を抑止すること²³⁾なども、文分割にとって必須の技術

であり今後の課題である。

分割文生成全体としては、分割パターンの不備による分割文生成の失敗が少ないこと、非対象文を除けば、ルール作成文で約 85%，それ以外で約 80% の文に対して有効な分割文を生成できることから、4 章で提案した分割文生成方式が有効であることがわかる。

6. おわりに

本論文では、接続構造解析に基づく複文の分割方式を提案した。本方式の特徴は、接続形式の導く従属句の独立性に着目することによって接続構造を解析し、分割点を特定すること、連用中止表現をその表現形態に着目することによって分類し、接続構造の解析および分割文の生成に必須となる意味用法の特定に利用することである。これらによって、文全体の構文解析を行わずに分割点を特定することができるとともに、原因や方法といった意味用法を担う連用中止表現を含む文に対しても、適切な分割文を生成することができる。本論文で提案した文分割方式は、推敲支援技術としてだけでなく、従来の構文解析技術と融合させることによって、日本文解析技術の高度化に寄与するものだと考えられる。

また、評価実験を行い、約 87% の文に対して、正しく接続構造を解析できること、約 86% の文に対して、正しく分割点を特定できること、約 93% の連用中止表現に対して、2 個以内の意味用法に限定できること、約 83% の文に対して有効な分割文を生成できることを示した。

本方式を拡張し、接続構造解析、分割文生成の精度を向上させるためには、マニュアル文章における述語の使用場面や意味的な情報を導入することによって、述語間の関係を詳細に解析する技術が必要である。今後は、5 章で考察した結果に基づき、接続構造解析や分割文生成を詳細化し、精度の向上を図る予定である。また、文章構造の解析や文章生成の研究^{19)~23)}を参考にして、本論文では検討の対象としなかった主題表現の連鎖や指示表現などの文脈的な要因が、接続構造、分割の適切性、分割パターンにどのような制約を与えるかについても検討していきたい。

謝辞 日頃ご指導いただく NTT 情報通信網研究所メッセージシステム研究部坂井陽一郎、坂間保雄主幹研究員、小原永主幹研究員、高木伸一郎主任研究員ほか、同研究部の方々に深く感謝いたします。

参考文献

- 1) 信国佳之：自然言語における長文分割方式，第 39 回情報処理学会全国大会論文集，2 F-1 (1990)。
- 2) 鈴木克志，高山泰博，太細 孝：標準文章処理システム，第 36 回情報処理学会全国大会論文集，5 U-1 (1988)。
- 3) 高橋善文，吉田哲三：計算機マニュアル推敲・査読システム MAPLE の開発と運用，情報処理学会論文誌，Vol. 31, No. 7, pp. 1051-1062 (1990)。
- 4) 林 良彦，菊井玄一郎：日本文推敲支援システムにおける書換え支援機能の実現方式，情報処理学会論文誌，Vol. 32, No. 8, pp. 962-970 (1991)。
- 5) 高橋昭男：わかりやすいマニュアルの作成法，日経 BP 社 (1985)。
- 6) テクニカルコミュニケーション研究会(編)：ユーザマニュアルのための文章・用字用語スタイルブック，日経 BP 社 (1988)。
- 7) 寺村秀夫：日本語の文法，国立国語研究所 (1980)。
- 8) 三上 章：現代語法新説，くろしお出版 (1972)。
- 9) 南不二男：現代日本語の構造，大修館書店 (1974)。
- 10) 益岡隆志：モダリティの構造と疑問・否定のスコップ，日本語のモダリティ，pp. 193-210，くろしお出版 (1989)。
- 11) 野田尚史：真性モダリティをもたない文，日本語のモダリティ，pp. 131-157，くろしお出版 (1989)。
- 12) 亀井真一郎，村木一至：Lexical Discourse Grammar の提案，電子情報通信学会自然言語とコミュニケーション研究会資料，86-7 (1986)。
- 13) 平井 誠，北橋忠宏：事象の事実性を用いた日本語の複文解析，情報処理学会自然言語処理研究会資料，63-5 (1987)。
- 14) 水野順子，中垣寿平：日本語の文の構造に関する一考察，情報処理学会自然言語処理研究会資料，76-4 (1990)。
- 15) 長尾 真，辻井潤一，田中伸佳，石川雅彦：科学技術論文における並列句とその解析，情報処理学会自然言語処理研究会資料，36-4 (1983)。
- 16) 北原保雄：日本語助動詞の研究，大修館書店 (1981)。
- 17) 国立国語研究所(編)：現代語の助詞・助動詞，秀英出版 (1951)。
- 18) 永野 賢：文章論総説，朝倉書店 (1986)。
- 19) 福本淳一：筆者の主張に基づく日本語文章の構造化，情報処理学会自然言語処理研究会資料，78-15 (1990)。
- 20) 小野顕司，浮田輝彦，天野真家：文脈構造の分析，情報処理学会自然言語処理研究会資料，70-2 (1989)。
- 21) 小野顕司，住田一男，浮田輝彦，天野真家：ラムダ記法による文脈構造の表現と処理，情報処理

学会自然言語処理研究会資料, 79-5 (1990).

- 22) 吉本 啓: 談話処理における日本語ゼロ代名詞の扱いについて, 情報処理学会自然言語処理研究会資料, 56-4 (1986).

- 23) 小谷 亮, 今村 誠, 近藤省造: 文章の機能的構造からの接続表現生成, 情報処理学会自然言語処理研究会資料, 82-7 (1991).

(平成3年9月30日受付)

(平成4年2月14日採録)



武石 英二 (正会員)

1963 年生. 1986 年早稲田大学理工学部数学科卒業. 1988 年同大学院理工学研究科数学専攻博士前期課程修了. 同年日本電信電話(株)入社. 現在, NTT 情報通信網研究所において, 日本文推敲支援システム等の自然言語処理システムの研究開発に従事. 人工知能学会会員.



林 良彦 (正会員)

1959 年生. 1981 年早稲田大学理工学部電気工学科卒業. 1983 年同大学院理工学研究科博士前期課程修了. 同年日本電信電話公社入社. 現在, NTT 情報通信網研究所メッセージシステム研究部主任研究員. 自然言語処理の研究に従事. 人工知能学会会員.