

## ヘルプテキストを対象とした修辞構造解析システムの実装と改良

安達 昌吾<sup>†</sup>杉本 徹<sup>‡</sup>芝浦工業大学大学院 理工学研究科<sup>†</sup>芝浦工業大学 工学部<sup>‡</sup>

## 1. 研究の背景と目的

家電製品の取扱説明書やソフトウェア付属のヘルプシステムについて、記述がわかりにくいという不満を感じた経験を持つ人は多くいる。この問題を解決するために、取扱説明書のテキスト構造を自動的に理解することで、ユーザの問い合わせに応じて必要な箇所のみを抽出し、文言をわかりやすく言い換えることのできる知的ヘルプシステムの開発に取り組んでいる。

この研究では、知的ヘルプシステムに用いるテキスト構造の自動解析システムの実装と改良を行う。人手によるテキスト構造の分析結果から得られたルールに基づいて実装したシステムを用いて、未分析のテキスト構造を解析することで、ルールの修正と拡張を行う。

## 2. 修辞構造解析システムの実装

## 2. 1. システムの概要

本研究では、伊藤らの提案したアルゴリズム[1]を基に修辞構造解析システムを実装する。このシステムは、以下の手順でヘルプテキストの構造を解析する。

(1) 入力されたヘルプページの HTML 解析を行い、文章を抽出する。

(2) 形態素解析を行い、文章を形態素列に分割する。

(3) 係り受け解析を行い、文節の係り受け関係を求める。

(4) 節間の修辞構造解析を行い、文を節に分割し、隣接した節と節の間の修辞関係の推定を繰り返し、解析木を生成する。

(5) 文間の修辞構造解析を行い、隣接した文と文の間の修辞関係の推定を繰り返し、解析木を生成する。

解析結果の例を図1に示す。

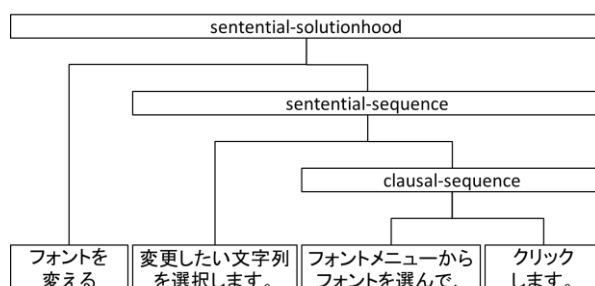


図1：システムの解析結果の例

## 2. 2. 節間の修辞構造解析

節間の修辞構造解析では、以下のルールを順に適用して、修辞関係の推定を行う。

(1) 左の節が特定の言語表現を含む場合、その言語表現に対応した修辞関係とする。

(2) 左右の節の動詞が特定のペアである場合、そのペアに対応した修辞関係とする。

(3) 隣接する節間の関係が同じでそれらが「手段」を表す「means」である場合、修辞関係は「連言」を表す「conjunction」とする。

(4) (1) ~ (3) のいずれにも該当しなかった場合、修辞関係は「順序」を表す「sequence」とする。

## 2. 3. 文間の修辞構造解析

文間の修辞構造解析は、以下のルールを順に適用して、修辞関係の推定を行う。

(1) 左の子の左端の文が「目的」を表す「purpose」あるいは「条件」を表す「condition」を含み、右の子の左端の文が「purpose」を含む場合、修辞関係は「conjunction」とする。

(2) 右の子の左端の文が「purpose」あるいは「condition」を含む場合、修辞関係は「詳述」を表す「elaboration」とする。

(3) どちらかの文が特定の言語表現を含む場合、その言語表現に対応した修辞関係とする。

(4) (1) ~ (3) のいずれにも該当しなかった場合、修辞関係は「sequence」とする。

## 2. 4. 実装したシステムの評価

システムの評価には、伊藤らがアルゴリズムを考案する際に分析した Microsoft Office Word 2000 のヘルプテキストから 25 タイトル 118 文、

Implementation and Improvement of a Rhetorical Structure Analysis System for Instructional Texts

<sup>†</sup>Shogo Adachi Graduate School of Engineering and Science, Shibaura Institute of Technology

<sup>‡</sup>Toru Sugimoto College of Engineering, Shibaura Institute of Technology

および後に発売された Microsoft Office Word 2007 のヘルプテキスト[2]から 28 タイトル 142 文を用いた。

実験の結果を表 1 に示す。

表 1: 実装したシステムの修辞構造解析の結果

評価項目	2000	2007
適切な修辞関係が得られた節間	94.7% (72/76)	58.1% (36/62)
適切な修辞関係が得られた文間	98.9% (92/93)	93.9% (107/114)

Word 2007 における節間の修辞構造解析で誤りが比較的多かった。その主な原因は、以下のように分類される。

(i) 解析木の生成順序の誤り

実装したシステムでは、右端の節から左に向かって解析木を生成するが、Word 2007 の分析結果には、左端の節や中央の節が最も深い葉となる解析木も存在する。

(ii) 動詞ペアの不足

動詞ペアを用いなければ適切な解析のできない節間であるが、提案されたアルゴリズムにはそのペアが定義されていない。

(iii) 対比表現の不足

例えば、左の節の動詞が「追加する」、右の節の動詞が「追加しない」であった場合、修辞関係は「連言」を表す「conjunction」となる。Word 2000 の分析結果には、このように左右の節で意味的に反対となる表現が存在しない。

また、Word 2007 における文間の修辞構造解析では、箇条書きの構造を表す HTML タグを修辞関係の推定に利用していないことによる誤りが見られた。

### 3. 修辞構造解析システムの改良

#### 3.1. 改善手法の考案

本研究では、実験結果に基づき Word 2007 においても適切な修辞関係が得られるように、節間の修辞構造解析アルゴリズムの改良を行った。

(i) 節間の解析木の生成順序の見直し

Word 2000 と Word 2007 の分析結果を基に、修辞関係ごとに解析木の生成における優先度を設け、様々な修辞構造への対応を可能とする。

(ii) 動詞カテゴリを用いた推定の導入

従来の動詞ペアによる推定では、対応できないペアが出現した場合に、必ず不適切な結果となる。そのため、新たなヘルプテキストを解析するたびに、動詞ペアの拡充が求められる。

そこで、同様の機能を持つ複数の動詞を纏め、1つのカテゴリとして扱うことで、動詞ペアを用いなければ適切な解析を行えない節間への対応を容易にする。動詞ペアを用いて推定する修辞

関係に「means」がある。この修辞関係は、左の節の動詞が「ある動作」、右の節の動詞が「その動作による変化」を表す。ここで、ヘルプテキストに表れる動詞をあらかじめ「直接ユーザが行う操作を表す動詞」と「変化を表す動詞」のいずれかに分類しておき、推定に用いることで、動詞ペアとして定義していなくとも、「means」という修辞関係の推定が可能となる。

(iii) 対比表現を用いた推定の導入

左右の節の動詞が意味的に反対となる場合、「連言」を表す「conjunction」とする。

また、文間の修辞構造解析において、新たに HTML タグを用いた推定を導入する。左右の文が、「<ul type="disc">」でタグ付けされていた場合、修辞関係を「conjunction」とする。

#### 3.2. 改良したシステムの評価

システムの評価には、Word 2007 のヘルプテキストから 28 タイトル 142 文、Word 2010 のヘルプテキストから 34 タイトル 158 文を用いた。

実験の結果を表 2 に示す。

表 2: 改良したシステムの修辞構造解析の結果

評価項目	2007	2010
適切な修辞関係が得られた節間	95.2% (59/62)	84.4% (54/64)
適切な修辞関係が得られた文間	95.6% (109/114)	93.5% (116/124)

表 1 に比べ、Word 2007 における節間の修辞構造解析では、誤りが大幅に減少した。また、Word 2010 における節間の修辞構造解析では、動詞カテゴリに含まれる動詞の不足による誤りが見られたものの、表 1 の Word 2007 よりは良い解析結果が得られた。

### 4. 結論

ヘルプテキストを対象とした修辞構造解析システムを実装した。また、実験結果の分析に基づいて解析アルゴリズムの修正と拡張を行った。改良後のシステムは、新しいヘルプテキストにおける節間の修辞構造解析で 84.4%の精度が得られたことから、高い汎用性を持っていると考えられる。今後は、他のソフトウェアのヘルプテキストでの実験、分析も検討していきたい。

#### 参考文献

- [1] 伊藤紀子, 杉本徹, 岩下志乃, 小林一郎, 菅野道夫: “日本語ヘルプテキストの修辞構造分析と対話型ユーザ支援システムへの応用”, 機能言語学研究 Vol.4, pp.83-104, 2007-04
- [2] 伊藤紀子, 岩下志乃, 杉本徹, 小林一郎: “3種類日本語ヘルプテキストの修辞構造分析と比較”, Proceedings of JASFL 4, pp.69-78, 2010-04