

4 J-2

## 文間文脈と文内文脈を利用した 日本語複文における省略要素の復元

土井伸一

日本電気株式会社 C&amp;Cメディア研究所

e-mail: s-doi@ccm.cl.nec.co.jp

### 1 はじめに

一般に自然言語では、情報伝達の効率化を図るため、他の手段（共有知識を含む）で伝達可能な情報は明示しないという原則の下、多くの省略表現が使用されている。日本語では特にこの傾向が著しく、日本語解析システムにおいては省略要素の復元が大きな課題となってきた。最近ではセンタリング理論を応用した研究 [4, 8] が広く行われているが、対象は主に単文間のローカルな結束性である。そこで、日本語の複文における省略復元の精度を向上するために、用言の意味属性や接続語の情報を利用する手法 [6] や、これらの情報とセンタリング理論を組み合わせる手法 [7] が提唱されている。

筆者らもこれまでに、「複文中の用言間の接続語のレベル」と「述語の変化素性」に基づいて日本語複文・重文の文構造を推定する語彙文脈文法 (Lexical Discourse Grammar: 以下、LDG) [3] を提唱し、省略要素の復元にも応用を図っている [1]。LDG の特徴は、文中の語彙の持つ機能情報を利用して、構文・意味解析に先立って大局的な文構造を推定することにある。以下、始めに LDG の概要を説明し、続いて LDG に基づく日本語複文における省略要素の復元手法について述べる。

### 2 LDG: 語彙文脈文法

LDG は、文中の機能語彙に着目して長文の大局的な構造を推定する手法である。文内文脈構造を特徴づける接続助詞等の表現を収集しておき、それらの語彙情報を用いて文構造を推定する [3]。この語彙情報としては、

1. 接続助詞等の機能語が持つ文中での切れ目としての強さ (接続語のレベル)
2. 用言節の変化素性 (定義・属性・変化等) に基づく係り先の制約

の2点を用いている。

日本語の複文では、主文からの従属節の独立性・従属度にレベルがあることが指摘されている [5]。特に節と節を結ぶ接続語は、節の内部に含み得る要素や節間の相

対的な包含関係により、レベル分けすることができる。LDG では、このような複文中の用言間の接続語のレベルを、節の内部にどのような要素を含み得るかによって、6段階に設定している。基準を表1に示す。

このような接続語の分類にはこれまでも様々な研究があるが、これらと比して LDG は、日本語解析システムに「文内文脈解析部を形態素解析部の直後に位置付け、構文・意味解析に先立って文内文脈構造を推定して構造解析の曖昧性を減じる」という形で応用していることが特徴である。また、接続助詞以外の接続語にもレベルを設定するとともに、音声 (ポーズ) 情報との関連の調査 [2] も行って、有効性を確認している。

### 3 LDG に基づく省略復元

これまでに提唱された接続語の情報を利用して複文中の省略復元を行う手法 [6, 7] では、基本的に、まず複文を単純に用言ごとに単文に分割して、その後、単文間の省略補完関係に対する制約として接続語情報を用いている。これに対して本手法では、始めに接続語情報を用いて文の大局的構造を推定し、続いて、単文間の省略補完関係を認定する。以下、本手法における日本語複文に対する文構造推定と省略復元のアルゴリズムを述べる。なおここでは、簡単のため、復元する省略要素を「ハ格」「ガ格」で表される主語に限定して説明する。

#### 3.1 文構造推定アルゴリズム

- 1) 節の右端の推定 文中の用言を全て抽出し、附属している接続語に対してレベル情報を付与する。
- 2) 節の左端の推定 各格要素、副詞等は、基本的に最も近い用言に係るものとする。ただし、
  - 2.1) ガ格 レベル5の接続語を伴う従属節は独自のガ格を持つことができないので、レベル5以外で最も近い用言に係るものとする。
  - 2.2) ハ格 他のハ格 (対照のハ) 等、ハ格のスコップを遮る要素がない限り、最も遠い用言 (基本的には文末) に係るものとする。ただし、一文一格、非交差等の条件によっては、レベル0と1の節は独自のハ格をとれる。
- 3) 節間関係の推定 接続語のレベル情報を利用して節間の包含関係を推定する。

Japanese Zero Pronoun Resolution Based on Inter-sentential Discourse and Lexical Discourse Grammar

Shinichi DOI

C&C Media Research Labs., NEC Corporation

表1 LDGでの接続語のレベル

レベル	レベルの定義基準	接続語の例	
Lev.0	あらゆる要素を従えられる	と(引用)	上位 ↑
Lev.1	終助詞がとれない	から、ので、けれども、ながら(逆接)	
Lev.2	断定のモーダルがとれない	れば、なら、ほか、場合、時、と(場合)	↓ 下位
Lev.3	推量のモーダルがとれない	前、あと	
Lev.4	時制がとれない	とともに、と同時に、限り、(て)	
Lev.5	従属節独自のガ格をとれない	ことなく、たまたま、つつ、ながら(順接)	

### 3.2 各レベルの節の特徴と省略復元アルゴリズム

#### 3.2.1 レベル0

引用節や連体節等の完全に独立した文であり、省略された主語は基本的には文外の要素から補完される。ただし、用言の中には、意味的に補文の主語を外側から制御するものが存在する。例えば、話者の意志を表す“～と決める”の場合は補文の主語が主文の主語と一致するし、“～と思う”の場合は補文と主文の主語が異なるのが原則である。なお現在のアルゴリズムでは、ハ格が一つしか存在しない場合は2.2)により主文に係るとしているが、文脈によっては従属節に係ることもありえるため、文間文脈知識も利用した優先度計算が必要である。

#### 3.2.2 レベル1

終助詞はとれないが従属節独自のハ格を持つことができ、独立した文に近い。従属節中の主語が省略されている場合に係り先の節と共有するか否かは状況に応じて意味的・語用論的に決まるが、基本的には共有の可能性が高い。なお、従属節が独自のガ格を持つ場合には、係り先の節とは主語が異なるのが原則である。

#### 3.2.3 レベル2～4

従属節独自のハ格を持つことはできないが、主語が省略されている場合の処置はレベル1と同様。

#### 3.2.4 レベル5

2.1),2.2)により、従属節は独自のガ格、ハ格を持つことはない。この従属節の(意味上の)主語は、係り先の節の主語と必ず一致する。

### 3.3 解析例

本手法による複文の構造推定、省略復元の例を示す。

- 1) 彼が笑いながら尋ねたので私は答えた。
  - 2) 彼が尋ねたので笑いながら私は答えた。
- ここでは、接続語“ので”と“ながら”の比較により、いずれの文もまず“ので”の直後で大きく切れて

1) [彼が[笑いながら]尋ねたので][私は答えた]。

2) [彼が尋ねたので][笑いながら][私は答えた]]。

という構造を持つことが推定される。これにより、「笑いながら」の主語は、前者では「尋ねたので」の主語である「彼」、後者では「答えた」の主語である「私」と認定することができる。

## 4 おわりに

語彙文脈文法に基づく日本語複文における省略要素の復元手法の概要を述べた。接続語等の語彙情報を用いて推定した文構造に基づいて、単文間の省略補完関係を認定することが特徴である。現在、コーパス上での評価、従来手法との比較を行っている。また、本稿では文内文脈を利用する手法について述べたが、センタリング理論等の文間文脈に関する理論との融合も検討中である。

## 参考文献

- [1] 土井 et al. “語彙文脈文法 (Lexical-Discourse-Grammar) とその大局的係り受け決定への応用 (II)”, 信学会言語理解とコミュニケーション研究会, NLC91-29, 1991
- [2] 土井 et al. “機能語に着目した文内文脈構造抽出手法 (LDG) と音声出力の制御”, 人工知能学会言語・音声理解と対話処理研究会, SIG-SLUD-9301-4, 1993
- [3] 亀井, 村木 “Lexical Discourse Grammar の提案”, 信学会言語理解とコミュニケーション研究会, NLC86-7, 1986
- [4] Kameyama “A Property-sharing Constraint in Centering”, ACL-86, 1986
- [5] 南 “現代日本語の構造”, 大修館書店, 1974
- [6] 中岩, 池原 “語用論的・意味論的制約を用いた日本語ゼロ代名詞の文内照応解析”, 自然言語処理, Vol.3, No.4, 1996
- [7] Okumura, Tamura “Zero Pronoun Resolution in Japanese Discourse Based on Centering Theory”, COLING'96, 1996
- [8] Walker et al. “Japanese Discourse and the Process of Centering”, Computational Linguistics, Vol.20, No.2, 1994