

## ⑰ A Discourse Structure Analyzer for Japanese Text

K.Sumita\*, K.Ono, T.Chino, T.Ukita, S.Amano(東芝, 日本)

### 発表要旨

我々の目的は談話構造解析器を実現することである。談話構造解析の研究はこれまでも幾つかあるが、実用的なシステムはこれまで実現されていない。我々のアプローチは言語外的な知識を用いず、言語知識だけを用いる方法である。テキストの大域的な構造を認識するのに大域的な言語的知識を利用し、部分的な構造を決定するのに局所的言語知識を利用する。これらの言語知識を形式化した後、実際のテキストを用いてこれらのテキストと手続きを評価する。

テキストの談話的構造は文と文の部分関係で構成された二進木として表現される。部分関係を18カテゴリに分類した。

このシステムは5つの部分からなりたっている。

1. 前処理：テキスト中の各文を形態素解析と統語解析を行なう。繰り返し出現する表現や話題を表す句を引き出す。結合関係を検出し、結合列を構成する。
2. 分割化：テキストの大域的なセグメントを検出し、制限された結合列を生成する。
3. 候補の生成：談話構造の候補を生成する。
4. 候補のしぼり込み：Thinking-flow原則に反する候補を取り除く。
5. 優先度の判断：話題とその参照表現に基づいて優先解を求める。

実験は18の記事(114段落)を用いた。

### 質疑応答

質問：日本語の文は接続詞を省略するが、どのように解消するのか？

回答：我々は接続詞だけを使うのではない。述語のパターンと単語の共起も使う。

質問：実験では1つのパラグラフ当たり何文あったか？

回答：3文から4文。

質問：1パラグラフの文の数が増えると計算量はどうなるか？

回答：文が14個の場合で候補は約70万になった。しかし、分割ルールを用いると、その平方根になる。

質問：あなたは同じ構造を機械翻訳で用いることが可能だと思いますか？

回答：はい。