

状況適応モデルによる 会話文理解実験システム

I C O T

第6研究室
株式会社 東芝

浮田 輝彦

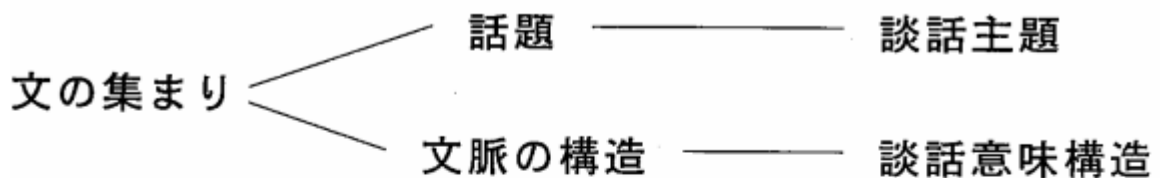
研究の目的

- 自然言語の複数の文を扱う文脈処理技術を開発する
- 文脈構造を抽出する方法を確立する
- 文脈処理モデルを開発する

研究経過

- 昭和60年度～昭和63年度
「意味談話理解のための知識表現実験システム」
対象分野の知識と文脈から会話文を理解する方式の検討
分野を限定した小規模な実験システムを試作
指示表現や省略表現の情報を推定
- 平成元年度
「状況適応モデルによる会話文理解実験システム」
文章の構造化処理の枠組みの検討
文脈構造の設計と抽出方法

文脈情報



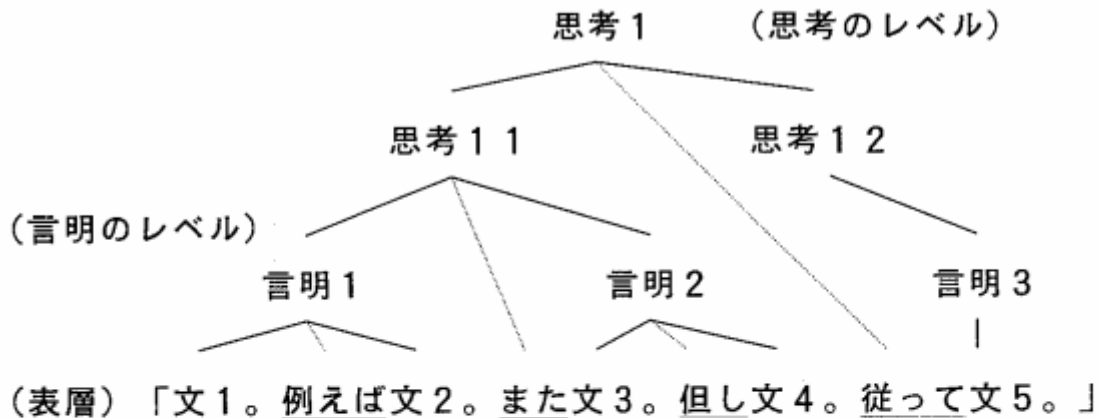
{言語学的知見}

{発話状況}

{世界知識}

{理解の原則・戦略}

文脈構造の表現

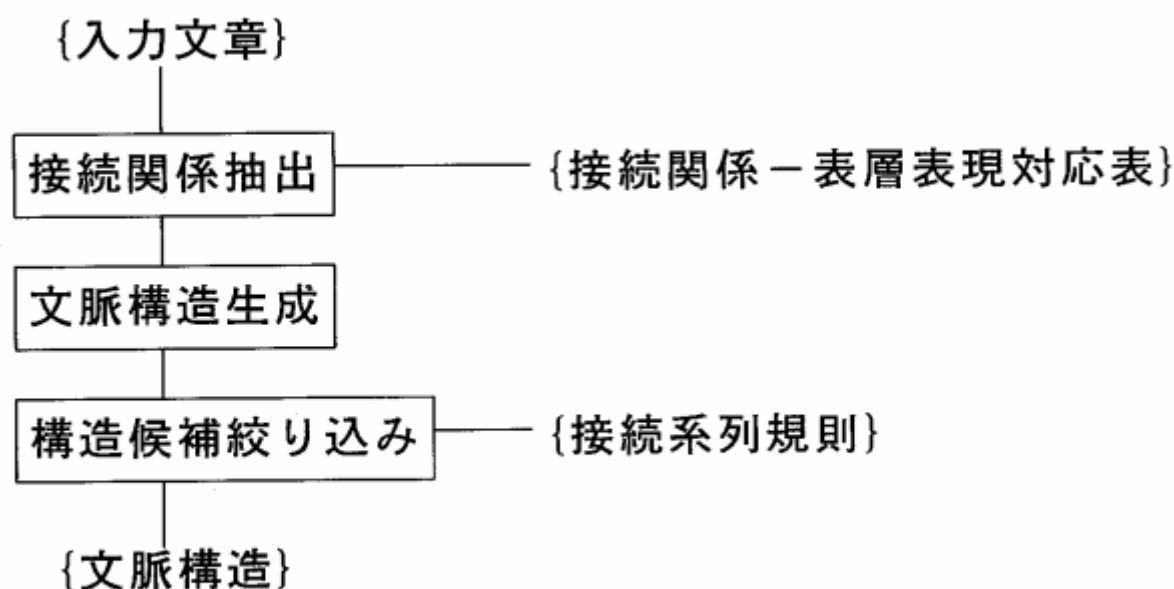


(((文 1 例えば 文 2) また (文 3 但し 文 4)) 従って 文 5)

文間の接続関係

例示	(例えば)	並列型：並列	(また)
重複	(というのは)	対比	(一方)
理由	(なぜなら)	直列型：順接	(従って)
補足	(但し)	逆接	(しかし)
背景	(従来)	同列型：同列	(すなわち)
定義	(…とは)	指示型：指示	(以下に…述べる)
ハイライト	(…こそ)	参照	(図 2 に…示す)
継続	(この)	転換	(さて)
		概括	(結局)
		解答	(その答は)

文脈構造抽出処理



接続系列規則の例

次のパターンを棄却する

- … 文1 直列型関係 (文2 直列型関係 文3 …
(だから) (だから)
- … 文1 同列型関係 (文2 直列型関係 文3 …
(すなわち) (だから)
- … 文1 補足) 文2 理由 文3 …
(但し) (なぜなら)

「文1 すなわち (同列型関係) 文2 だから (直列型関係) 文3」

○ ((文1 すなわち 文2) だから 文3)

X (文1 すなわち (文2 だから 文3))

処理の例（１） <入力文章からの接続関係抽出>

①文書はユーザの考えを言語化、ほかの人に理解してもらうために作られる。②そのために文書の論理的展開をより明確にするために、章や節のまとまりに分割し、見出しを付けることにより、わかりやすい文書にする。③さらに見出しには番号や記号を付れたり、また字体や文字サイズを変えたりする。④またタブ、インデント、センタリングなどのレイアウトを行い紙面上の配置を考えながら文書を作成する。⑤例えば、学会論文などは、標題、著者名、所属は紙面の上部に1段組で、章、節などの本文は2段組で割り付けるものがある。⑥このように文書のレイアウトは論理的な構造に依存している。⑦これらのレイアウト作業を自動化するためには、文書の論理構造の抽出が不可欠である。

処理の例（２）

<構造候補生成>

(((((1 順接 (2 並列 3)) 並列 4) 例示 5) 概括 6) 継続 7))	○
(((((1 順接 2) 並列 3 並列 4) 例示 5) 概括 6) 継続 7))	○
((((1 順接 (2 並列 3)) 並列 (4 例示 5)) 概括 6) 継続 7))	○
((((1 順接 (2 並列 3)) 並列 4) 例示 (5 概括 6)) 継続 7))	X
((((1 順接 (2 並列 3)) 並列 4) 例示 5) 概括 (6 継続 7))	X
((((1 順接 (2 並列 3 並列 4)) 例示 5) 概括 6) 継続 7))	○
.....	

(90種類)

<接続規則による判定>

(((((1 順接 (2 並列 3)) 並列 4) 例示 5) 概括 6) 継続 7))	
(((((1 順接 2) 並列 3 並列 4) 例示 5) 概括 6) 継続 7))	
((((1 順接 (2 並列 3)) 並列 (4 例示 5)) 概括 6) 継続 7))	
((((1 順接 (2 並列 3 並列 4)) 例示 5) 概括 6) 継続 7))	
.....	

(8種類)

試作システムの処理例

The screenshot displays a multi-window application interface for document structure analysis. At the top, there are control windows for 'execute.select', 'menu.window', 'select.move', and 'input.window'. The main workspace is divided into several panes:

- cond_window:** Displays a tree diagram titled '文脈構造' (Context Structure) with nodes labeled '1' through '7' and arrows indicating relationships.
- structure_candidate_window:** Displays a list of candidate structures titled '文脈構造候補' (Context Structure Candidates), showing hierarchical paths of nodes like '1 ((1, MR, 2), MR 3, MR 4), MR 5, MR 7 0'.
- diag_window:** Shows a table with columns for 'オーダリング規則', '抽出規則1', '抽出規則2', and '抽出規則', along with their respective parameters.
- source_text_window:** Displays a table with columns 'ch', 'sp', 'op', 'div' and corresponding text analysis results.

まとめ

- 自然言語文章の文脈構造記述の設計
- 文脈構造抽出方法の基礎的検討
- P S I 上に実験システムを試作
 - 文脈構造候補数を $1 / 10 \sim 1 / 20$ に削減

今後の課題

- 言語的手掛かりの利用による文脈構造抽出方法改良
 - 同一語の反復利用, 文述語のタイプ
- 対象分野知識を用いる文脈処理モデルへの展開