

TM-0415

要約支援システムCOGITO

— 文書の構造解析 —

小松英二，加藤安彦，安原 宏，椎野 努

November, 1987

©1987, ICOT

ICOT

Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03) 456-3191~5
Telex ICOT J32964

Institute for New Generation Computer Technology

要約支援システム COGITO

— 文書の構造解析 —

小松 英二 加藤 安彦 安原 宏 椎野 努
沖電気工業（株） 総合システム研究所

要約支援システムCOGITOは、文脈処理として、文書の構造を決定する。文書の構造解析には多くの要素に着目した解析が必要である。このため、本システムでは、複数の方法により決定される結果をそのまま結びつけて文書の構造とし、同時に、全体の解析率を高める方法を取った。それぞれの方法としては、当面は言語的制約と内容の関連にしばっている。対象とする文書は、情報処理機器関連の文書である。文書の構造の単位としては、単文を用いており、前処理としては、簡単なセマンティック・パージングと照応処理を行う。

Summarization Support System, COGITO

— Structure analysis of text —

Eiji Komatsu Yasuhiko Kato Hiroshi Yasuhara Tautomu Shiino
Systems Laboratory Oki Electric Industry Co., Ltd.
4-11-22 Shibaura Minato-ku Tokyo

Summarization support system COGITO decides the structure of text as contextual analysis. Analyzing text structure requires consideration of many elements. And to increase the successful rate of whole analyzing, we choose the approach such as combining all results each decided by a method of different paradigm. As individual method, at present, we select methods using linguistic restrictions and contentual relations. Target text is those related to informational processing equipments. Unit of text structure is simple sentence. In advance to the processing, a simple semantic parsing and an anaphoric processing are performed.

1. はじめに

現在、我々は要約支援システム COGITO ([1]) を開発中である。要約の方法は、大きく分けると

- (1) キーワードの頻度を用いて重要な文を決定する ([2])
- (2) 文と文の関係を用いて重要な文を決定する ([3]、[4])
- (3) 何らかの理解を行って重要な文を決定する ([5])

などの方法がある。本システムでは、要約以外のアプリケーションにも適用可能な方式として、(3) のアプローチを取っている。この際、理解として、文の構造、照応を含めた文脈処理を行う。文書の構造の決定方法には、因果関係による方法 ([6])、物語文法による方法 ([7]) 等があるが、日常生活で通常現われる文書の複雑さを考えると、言語的制約・意図・内容の関連・因果関係等、表層レベルから深いレベルまで多くの要素に着目する必要があると思われる。このため、本システムでは、複数の方法によって決定される結果をそのまま組み合わせる文書の構造とすることにより、全体の解析率を高めるアプローチを取った。それぞれの方法については、意図や因果関係等の深い解析は当面避け、言語的制約と内容の関連等、表層に結びつきやすく、深い解析に必要な要素にしばった。本稿では、情報処理機器関連の新聞記事を対象にした文書の構造解析の方法について報告する。

2. 意味表現

本システムでは、単文を文書の構造を決定するための基本単位とする。単文は、用言を1つだけでもつように文書を分割する。図1に、単文の分割例を示す。同じ番号の文は、一続きの単文と考える。また、文書構造の表現の例を図2に示す(図1と2は、共に机上で解析したもの)。図中の数字は上述した単文を表す。図2の表現を文書構造情報と呼ぶことにする。文書構造情報は4種類の解析処理によりほぼ独立に決定された依存関係をまとめて表しているため、木構造ではなく、ネットワークになる。図中のアークの矢印は、付随的な文から主要な文に向かってつけられており、両者が同等の場合には矢印をつけていない。表1に単文の依存関係の関係名の一覧を示す。関係名は、どの構造決定処理により決定されたかにより4つにグループ分けしてある。関係の詳細は、5節で述べる。

1. A社、B社は
2. 今後の主力製品として期待されている
1. 32ビットMPUでオリジナル路線の方針を固めた。
3. 32ビットハイエンドプロセッサでは、開発投資、ソフトウェア資産とアプリケーションの拡がりや影響力などから従来のような安易な対応には問題がある
4. とし、
5. 国内半導体メーカーでは、
6. セカンドソース、オリジナル路線と平行して
5. 慎重な選択を行ってきたが、
7. MPUも多様化、個性化の時代を迎え、
8. 独自開発品を持つ
9. ことが重要との判断から、
10. オリジナル路線でいく
11. ことを決めた。
12. 国内メーカーでは、C社がオリジナルMPU「Xシリーズ」で32ビットMPUの販売を開始しており、
13. 両社の32ビットMPUのオリジナル路線はこれにつづくもの。

図1 単文の分割例

本研究は第5世代コンピュータプロジェクトの一環としてICOTからの委託で行われたものである。

ト値に対応)に分類して辞書に記述している。

意味解析では、対象フレームを生成するために、動詞概念辞書と製品知識の2つの知識を用いている。図5に動詞概念辞書の例を示す。動詞概念辞書は、動詞の格に当てはまる対象の候補及び動詞のある格に対する他の格の関係を記述している。また、図6は製品知識の例である。製品知識は、製品に関する単語レベルの詳細なシソーラスとpart_of等を含む対象モデルを組み合わせた知識である。動詞概念辞書の属性名及び属性値は、製品知識の属性名、属性値と矛盾しないようなチェックを行なう。

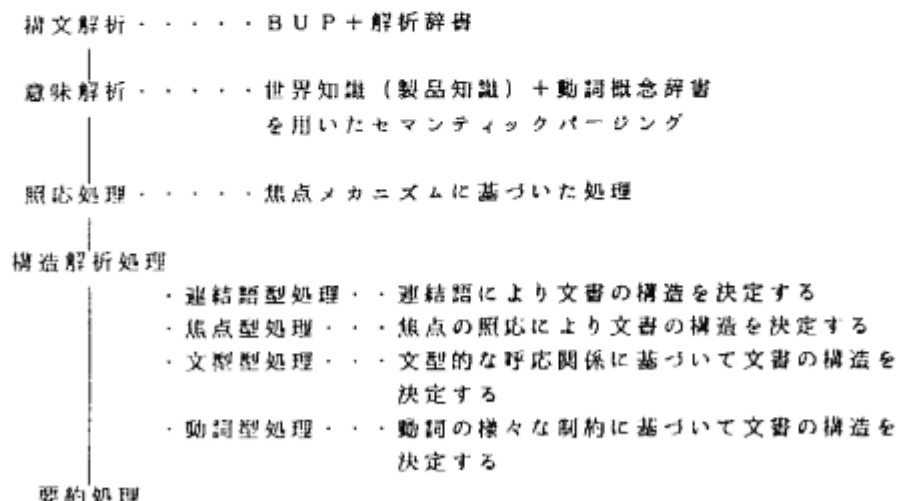


図3 要約支援システムの概略の処理フロー

名詞句: 3 2 ビットスーパーパーソナルコンピュータ i f 1 0 0 0

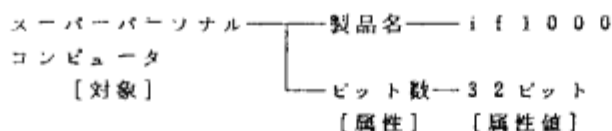


図4 対象フレームの例

見出し: 販売する

```
agent : {[org.*company_name].会社}, . . . . . 対象の条件
agent: [*product_company.会社名]
obj   : [cpro.販売製品名]          属性の条件
recip : [hum.販売対象]
action: [{].action}

obj   : {[cpro].製品},
agent: [*product_company.販売会社名]
obj   : [cpro.製品名]
recip: [org.販売対象]
action: [{].action}

recip: {[org.*company_name].会社}
recip: {[hum].消費者}
```

*印は、条件チェック用の手続名を表す

図5 動詞概念辞書の例

図7に単文の用言が他動詞場合と「ダ」文の場合の対象フレームの生成例を示す。フレームに他の格の情報を取入れ、照応の際の一致のチェックを正確にするのが特徴である。

4-2 照応処理

意味処理で生成された対象フレームは、照応処理で、既に生成されているフレームとマッチングが行われ、マッチングが成功したときは、同じ識別子をつけ、以下同一のフレームと考慮して処理を行う。図8にフレームのマッチング・ルール及び例を示す。照応処理により、フレームの対象がさらに詳細化される可能性がある。なお、照応の候補となるフレームの選択の順序については、焦点(Sidner, [7])に基づいている。

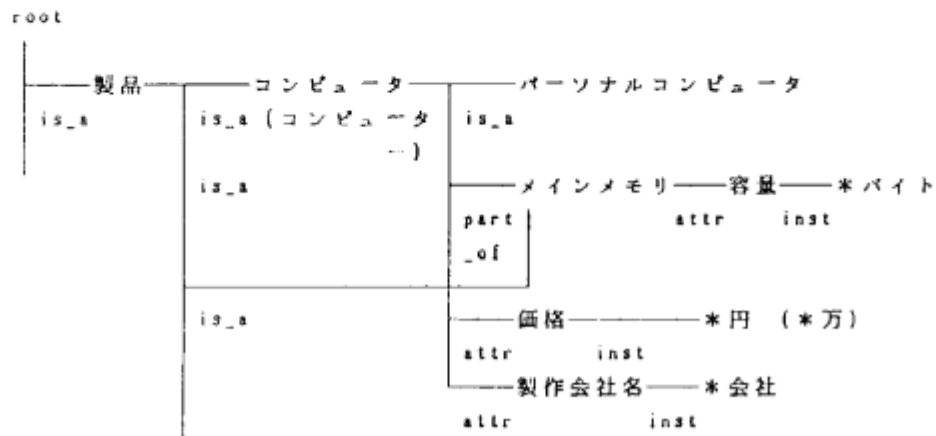
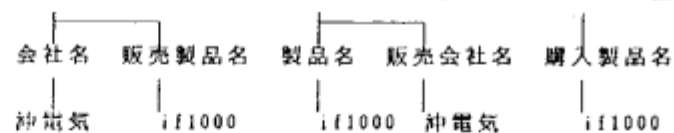


図6 製品知識の例

〔入力文〕 沖電気はパーソナルコンピュータif1000を販売した。(他動詞の場合)

〔対象フレーム〕 会社は パーソナルコンピュータを rootに 販売した。



〔入力文〕 コンピュータのCPUはMC68000だ。(「ダ」文の場合)

〔対象フレーム〕 コンピュータ——CPU——CPU名——MC68000
part_of attr inst

図7 対象フレーム生成の生成例

〔フレームのマッチングのルール〕

対象同士・・・一致、上位一下位、対象一属性等 → 成功
属性値同士・・・矛盾がない

〔成功例〕

製品——価格——500万 <——> コンピュータ

〔失敗例〕

製品——容量——5メガ <——> ハードディスク——容量——20メガ

図8 マッチング・ルール及び例

5. 文書の構造解析

この節では、文書の構造解析を行う処理について述べる。構造解析では、いくつかの処理が独立に文書を解決するが、必要に応じて他の処理の結果を用いることも行う。

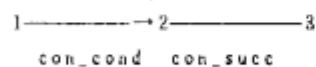
5-1. 連結語型処理

接続詞・接続助詞・接続詞的名詞等（連結語と呼ぶ）を用いて単文同士の依存関係を決定する。連結語は、関係名に対応して分類しておく。単文を、連結語により、条件文と結論文に分類する。条件文は、最も近い結論文に結び付くことにする。

例：〔入力文〕

1. 50個の専用超LSIを採用することにより、（条件文）
2. カラー三次元ソリッドイメージ描画が1秒以内で、（結論文）
3. 同等機に比べ10～20倍の高速処理を実現した。（結論文）

〔文書の構造〕



5-2. 焦点型処理

本システムでは、照応処理において、焦点を用いているが、ここでは、文書の構造決定に関係ある処理をまとめて、焦点型処理としている。焦点型処理では、まず文の焦点を決定する。焦点は、文の主語に当る概念で、Sidner ([7]) に基づいている。焦点の決定は、次のようにして行う。

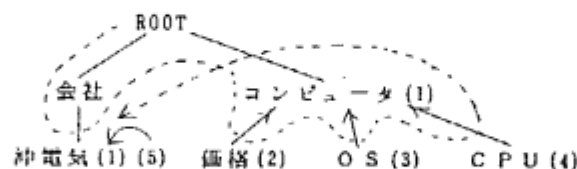
- 1). 「は」のついている格があるときは、その格の対象フレームを焦点にする。
- 2). 「だ」文の場合には、最も下位の対象フレームを焦点にする。
- 3). 1), 2) 以外は、以前のフレームを参照している最も先頭の格の対象フレームを焦点にする。
- 4). 1)～3) 以外は、最も先頭の格の対象フレームを焦点にする。

焦点は、製品知識中で下位に移動するときは、焦点スタックにプッシュし、以前の焦点に戻ったときにポップする。従って、焦点スタックには、現在の文の焦点の呼出し元である文と対象フレームが記憶されている。焦点の照応関係により文と文の関係を決定する。他の処理との関連として、連結語型処理の条件文に対しては、焦点スタックの操作を行わないことにしている。

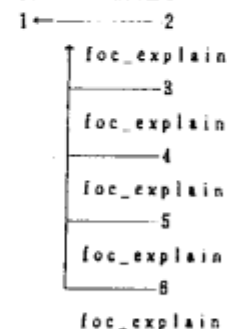
例：〔入力文〕

1. 社電気はスーパーパーソナルコンピュータif1000を開発した。
2. 価格は250万円。
3. OSにIF-UXを採用し、
4. CPUにMC68000を用いている。
5. 同社は、...

〔焦点の移動と単文の依存関係〕



〔文書の構造〕



ト値に対応)に分類して辞書に記述している。

意味解析では、対象フレームを生成するために、動詞概念辞書と製品知識の2つの知識を用いている。図5に動詞概念辞書の例を示す。動詞概念辞書は、動詞の格に当てはまる対象の候補及び動詞のある格に対する他の格の関係を記述している。また、図6は製品知識の例である。製品知識は、製品に関する単語レベルの詳細なシソーラスとpart_of等を含む対象モデルを組み合わせた知識である。動詞概念辞書の属性名及び属性値は、製品知識の属性名、属性値と矛盾しないようなチェックを行なう。

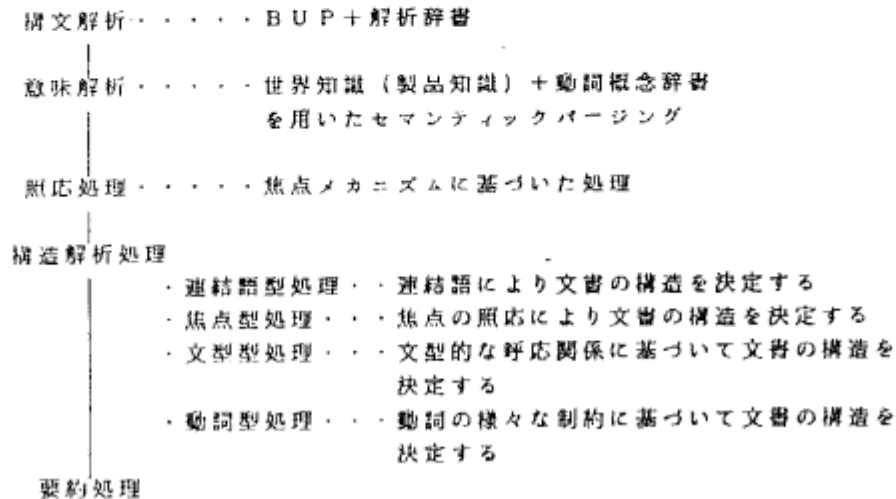


図3 要約支援システムの概略の処理フロー

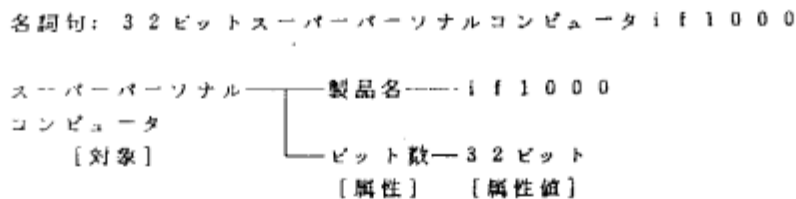


図4 対象フレームの例

見出し: 販売する

```

agent: {[org.*company_name].会社}, . . . . . 対象の条件
      agent: [*product_company.会社名]
      obj  : [cpro.販売製品名]          属性の条件
      recip: [hum.販売対象]
      action: [()].action
obj    : {[cpro].製品},
      agent: [*product_company.販売会社名]
      obj  : [cpro.製品名]
      recip: [org.販売対象]
      action: [()].action
recip: {[org.*company_name].会社}
recip: {[hum].消費者}

```

*印は、条件チェック用の手続名を表す

図5 動詞概念辞書の例

対になる単語の判定により単語同士の依存関係を決定する。

1. ①
2. ②

1-----2
typ_parallel

用言に関する種々の制約 ([9]) を用いる解析を行う。現在は、2つの処理からなる。

- 例：「入力文」

1. 土地を売り、
.....、
2. その金でマンションを買った。

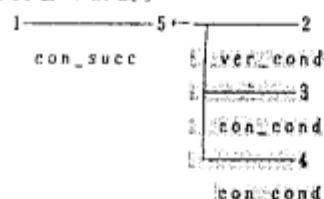
$$1 \leftarrow \text{-----} n$$

ver_ana

- 例：[入力文]

1. 国内半導体メーカーでは、慎重な選択を行ってきたが、(結論文)
2. MPUも多様化、個性化の時代を迎え、(結論文)
3. 独自製品を持つことが重要との判断から、(条件文)
4. オリジナル路線に決定した、(条件文)

[文書の構造]



この部分は他の処理による

本システムは、世界知識を多用した文書理解を目指しているが、同時に、fail tolerantなシステムも目指している。現段階では、言語的な知識と、製品のシソーラスを世界知識とした処理に制限しているが、今後、常識、意図、因果関係等の処理をつけ加えて性能向上を図る予定である。

本研究に当り、ICOT第2研究室の内田室長の御指導に感謝します。

- [1] 北. 小松. 安原: 要約支援システム COGITO. 自然言語処理研究会 58-7. 1986.
- [2] Luhn, H. P.: The Automatic Creation of Literature Abstracts. IBM J. Res. Dev. Vol2 p. 159. 1958
- [3] Hobbs, J R.: Coherence and Interpretation in English Texts. in IJCAI85.
- [4] 内海、重永: 英語文章の大意生成、情報処理学会研究報告、86-NL-54、1986年3月
- [5] Fun, D., Guda, G., Tasso, C.: Step toward the evaluation of text. in IJCAI.
- [6] Schank, R.: Understanding Paragraphs. Tech. Report 6. Istituto per gli studi semanticie cognitive. Castagnola, Switzerland. 1974.
- [7] Rumelhart, D: Notes on a Schema for Stories, in Bobrow & Collins. eds., Representation and Understanding. Academic Press. 1975.
- [8] Sidner, C. L.: Focusing in the Comprehension of Definite Anaphora. in Artificial Intelligence. The MIT Press. ss
- [9] 久野昭: 話話の文法