解　説

要約支援システム COGITO†

安原　宏††　小松英二††
日比　孝††　加藤安彦†††

1. はじめに

本稿では，自然言語理解システムの事例として，要約支援システム COGITO (COntext analysis and Gathering Important Text Objects) について述べる[1),2)]．コンピュータによる要約処理の方式を大別すると，文字列処理による表層的な要約と構文・意味処理による要約がある．後者はさらに，トップダウン的な要約とボトムアップ的な要約とに分けられる．表層的な要約では，単語の頻度[3)]・要約的表現・キーワード・テキスト内の位置などを用いて，重要箇所を決定する．トップダウン的な要約では，システム側で内容を予測し，抽出すべきデータをスクリプト[4),5)]や物語文法[6)]で記述しておき，これを用いて文書を解析することにより，重要箇所や重要事項を抽出する．ボトムアップ的な要約では，文書を談話構造などの意味表現に変換し，文書に即して重要箇所を決定する[7)-10)]．以上のように，要約にはさまざまな方法が提案されているが，これらをまとめると，表-1 に示すようになる．

表-1　要約の方法

方式		解析	→ 重要性評価	→ 生成
表層的解析による		●単語頻度解析など	●頻度などによる評価	●単語や文の抜粋
構文・意味解析による	トップダウン	●事項や事象のスクリプトへの埋め込み ●物語文法による文書の構造の解析	●重要項目の選択	●スクリプトからの文書生成
	ボトムアップ	●意味表現への変換	●重要な意味表現の選択	●意味表現からの文書生成

† Summarizing Support System COGITO by Hiroshi YASUHARA, Eiji KOMATSU, Takashi HIBI (Oki Electric Industry Co Ltd.) and Yasuhiko KATO (Japan Electronic Dictionary Research Institute Ltd.).

†† 沖電気工業(株)総合システム研究所
††† (株)日本電子化辞書研究所　本解説は沖電気工業(株)に在職中の成果に基づいてなされたものである．

一方，人間による要約に関しては，まず，斜め読みで文書の大筋を理解するような要約と詳細な理解をともなう要約がある．さらに，要約結果も，テキストの内容（論説文，説明文，新聞記事など）や利用者の目的（キーワード抽出，5W1H，テキストの概要）によって大きく異なる．COGITO は，製品記事を対象として開発したが，処理方式としては，できるだけ汎用をめざし，表-1 における方法を融合して用いることにより，人間の要約処理と類似した結果が得られることを目標としている．要約の出力形態としては，重要性の高い箇所（後述）の抜粋形式とする．

要約で第一に問題になるのは，重要性の定義である．Luhn[3)] は文書を特徴づけるような単語を重要語と定義するとともに，頻度により重要語を自動的に抽出する方法を提案した．同様の考え方を進めると，文書を特徴づける単文，文，段落（形式段落．以下，単に「段落」といえば形式段落を指す）の存在が考えられる．COGITO では，単語，単文，文，段落を総称して，重要性評価の単位になるという意味で，「要約ユニット」あるいは単に「ユニット」と呼び，文書を特徴づけるようなユニットの集まりを要約として定義する．重要性の高いユニットの判定方法として，以下のような条件を設定し，COGITO により検証を行うことにした．

① ［曖昧回避条件］
省略すると文書の意味が曖昧になるユニット．

② ［復元困難条件］
省略すると推論によって復元が難しいユニット．

③ ［意図伝達条件］
著者が読者に伝えようと意図しているユニット．

④ ［指示条件］
読者が特に指示したユニット．

曖昧回避条件は，他のユニットと格関係・照応関係・一貫性関係（後述）などで多く結びついているユニット．復元困難条件は，意外性や特異性のある事

象，数値，固有名詞などのユニット．意図伝達条件は，文書の作成された目的に深く関連するユニット（たとえば，新聞記事では5W1Hなど）．指示条件は，一般的に読者が関心をもつユニットや，個人的に興味をもっているユニット．上記の4つの条件は複合的な条件であり，一つの条件を満たすだけでは，重要ユニットとは限らない．COGITOでは，主に文書解析の意味表現を用いて曖昧回避条件の重要性評価を行い，表層的情報や浅い文書解析の結果を用いて他の三つの条件の重要性評価を行っている．

以下，本稿では曖昧回避条件を中心にして，2. 本システムの概要，3. 文脈処理方式，4. 重要性の評価方式と評価結果の出力，5. おわりについて述べる．

2. COGITO システム概要

図-1にCOGITOの概略構成図を示す．本システムは，要約プログラム，及び，要約を支援する要約エディタ・世界知識エディタからなる．言語はすべてESPを用いており，PSI-II上にインプリメントした．

（1） 要約プログラム

要約プログラムには，日本語を意味解析まで行い単語を概念化した中間表現テキストを入力とする．中間表現は図-2に示すように，格フレームを概念同士の2項関係とみて1階述語で記述したものである．中間表現の第1～3項は，(被修飾語の概念)—(格又は修飾関係)—(修飾語の概念)，及び，(概念)—(属性子)—(属性値）の2種類のフォーマットがある．第4項のリストは，要約エディタの画面で原文との対応をとるためのインデックスであり，段落番号，文番号，単文番号，第3項の概念の単語番号（第3項が属性値の場合は第1項）を表す．

（i） 文脈処理モジュール

文書解析としては，中間表現テキストから，照応関係・一貫性関係・段落関係（後述）を生成するための文脈処理を行う．また，文を超えた処理ではないが，重要性評価で用いるために，単文の分類として，事実か意見かの分類を決定する．

（ii） 重要性評価モジュール

表層的な重要性評価ルールと深層的な重要性評価ルールにより，ルールごとに各ユニットに対して重要性評価値を与え，それぞれのルールで決定した重要性評価値の重みつきの総和として，各ユニットの重要性

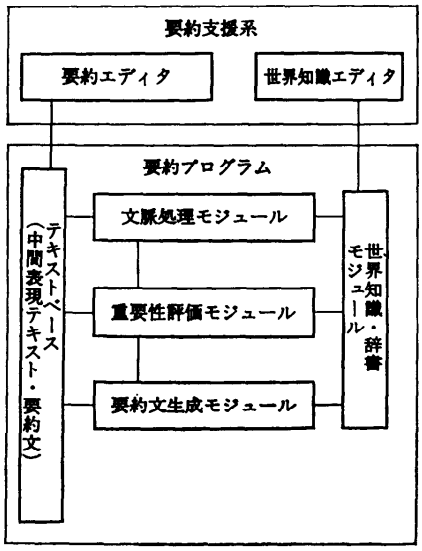


図-1 COGITOの概略構成図

評価値を算出する[11]．重要性評価値はユニット同士の相対的な重要性を示す数値で，プラスで数値が大きいほど重要であり，要約から意識的に排除するものにはマイナスを付与する．

（iii） 要約文生成モジュール

重要性評価値を用いて，重要ユニットの抜粋を行う．

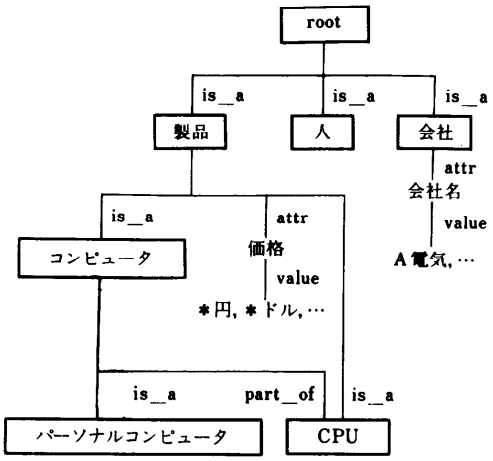
（iv） 世界知識・辞書モジュール

世界知識は図-3に示すように，複数の上位ノードをもつことを許した概念の階層構造である．概念は対象，属性，属性値に分類されており，対象と属性との間に対象-属性関係，属性と属性値との間に属性-属性

フォーマット：
meaning([被修飾語の概念, 識別子], 格又は修飾関係,
[修飾語の概念, 識別子],
[段落番号, 文番号, 単文番号, 修飾語の単語番号]).
meaning([被修飾語の概念, 識別子], 属性子, 属性子の値,
[段落番号, 文番号, 単文番号, 被修飾語の単語番号]).

例：
「A電気は18日1メガビットのRAMを生産すると発表した.」
単文1 単文2
meaning([発表する, 6], agent, [A電気, 1], [1, 1, 1, 1]).
meaning([発表する, 6], time, [[nil, 18, 日], 2], [1, 1, 1, 2]).
meaning([RAM, 4], [mod, def], [[nil, 1, メガビット], 3], [1, 1, 2, 3]).
meaning([生産する, 5], obj, [RAM, 4], [1, 1, 2, 4]).
meaning([生産する, 5], tense, present, [1, 1, 2, 5]).
meaning([発表する, 6], ref, [生産する, 5], [1, 1, 1, 6]).
meaning([発表する, 6], tense, past, [1, 1, 1, 7]).

図-2 中間表現のフォーマットと例



注) ●□のついたノードは対象を表し，その他のノードは属性・属性値を表す
●is_a：上位-下位，part_of：全体-部分，attr：対象-属性，value：属性-属性値
●＊は任意の数値を表す

図-3　世界知識の例

値関係，対象同士の間に上位-下位，全体-部分関係が定義されている．世界知識の構造は，照応のための名詞句の意味表現としても用いられており，本稿では，これらの関係が意味表現として用いられた場合には，対象-属性-属性値の組の集まりをフレーム，対象同士の関係をフレーム同士の関係，属性をスロット，属性値をスロット値と呼ぶ．

辞書は，単語の文法情報や単語に対応する世界知識の概念や文法情報を記入した解析辞書と，省略などを補う用言概念辞書からなる．

（2）　要約エディタ・世界知識エディタ

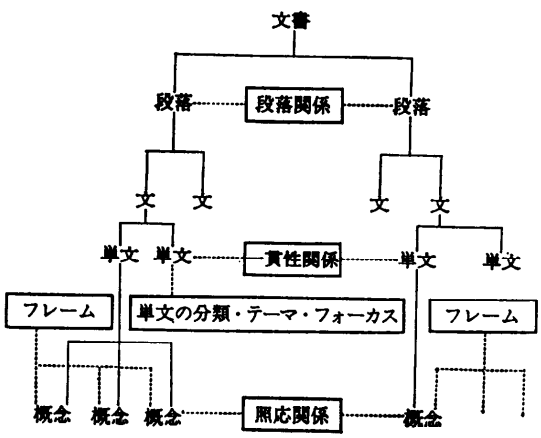
要約エディタは，画面に表示された原文と内部の中間表現の対応がついた自然言語用のエディタであり，原文における抜粋箇所を明確にするため，重要な部分のアンダーライン表示を行い，抜粋箇所及び抜粋による要約文のポストエディットを行う．

世界知識エディタは，世界知識の表示や，世界知識に未登録な概念の抽出・登録を行う．

これらのエディタは，要約システムのマンマシンインタフェース及び支援系として位置づけられる．

3. 文脈処理方式

要約システムの重要性評価においては，一文だけでなく前後の関係から重要性が決まる場合がある．たとえば，性質を述べた文でも，文書の中心的な事物に関する文とそうでない事物に関する文では，まったく重



注) □のないノード：中間表現の構造
□付ノード　　：文脈的な意味表現

図-4　意味表現の体系

処理能力最大のスーパ電算機88年末にも発売
AAAAAAAA社

米国のコンピュータメーカ，AAAAAAAA社のBBBBBB副社長は28日，現段階では世界で最も処理能力の大きい次世代スーパコンピュータ「CCC—5」の開発を進め，来年末から64年初めにかけて出荷する方針を明らかにした．

「CCC—5」の処理スピードは，ピーク時で16ギガFLOPS（1ギガFLOPSは，1秒間に10億回の小数計算をする能力）．現在，世界で販売されているスーパコンピュータで最も処理スピードの速い米国DDDDDDDDD社のEEE—10Eモデル（6.9ギガFLOPS）の2倍以上となる．また，AAAAAAAA社製品で最も処理スピードの速い「CCC—4」（1.95ギガFLOPS）の約8倍．記憶容量も5，6倍になるという．わが国で稼働中のスーパコンピュータは現在，約100台といわれ，うちAAAAAAAA社の設置台数は7台，受注済みを含めても11台．BBBBBB副社長は，米商務省の主催で27日から東京，大阪で商談セミナを行っている米国スーパコンピュータ貿易使節団の一員として来日，官公庁を中心に売り込みを続けている．

図-5　例文の原文

要性が異なる．このため，本システムでは重要性評価のための文脈処理を行う．

一般に要約は，個々の文ではなく，文書に対して適用されるものである．したがって，文書全体について処理が成功しなければならず，文書に定型的なフォーマットがある場合を除いては，多くの文について適用できる汎用性のあるボトムアップ的な解析方式が必要であると考えた．COGITOでは，文脈処理のためのトップダウン的な知識としては，2.で述べた，事物についての階層関係の知識だけを用い，接続語・埋め込み関係といったボトムアップで，かつ，比較的浅い処

1. 米国のコンピュータメーカ，ＡＡＡＡＡＡＡＡ社のＢＢＢＢＢＢ副社長は28日，方針を明らかにした．
2. 次世代スーパコンピュータ「ＣＣＣ—５」の開発を進め．
3. 現段階では世界で最も処理能力の大きい．
4. 来年末から64年初めにかけて出荷する．
5. 「ＣＣＣ—５」の処理スピードは，ピーク時で16ギガＦＬＯＰＳ（1ギガＦＬＯＰＳは，1秒間に10億回の小数計算をする能力）．
6. 米国ＤＤＤＤＤＤＤＤＤＤＤ社のＥＥＥ—10Ｅモデル（6.9ギガＦＬＯＰＳ）の2倍以上となる．
7. スーパコンピュータで最も処理スピードの速い．
8. 現在，世界で販売されている．
9. また，「ＣＣＣ—４」（1.95ギガＦＬＯＰＳ）の約8倍．
10. ＡＡＡＡＡＡＡＡ社製品で最も処理スピードの速い．
11. 記憶容量も5，6倍になるという．
12. スーパコンピュータは現在，約100台といわれ，
13. わが国で稼働中の
14. うちＡＡＡＡＡＡＡＡ社の設置台数は7台，
15. 受注済みを含めても
16. 11台．
17. ＢＢＢＢＢＢＢ副社長は，米国スーパコンピュータ貿易使節団の一員として来日，
18. 米商務省の主催で27日から東京，大阪で商談セミナを行っている．
19. 官公庁を中心に売り込みを続けている．

注）……：文の境界

図-6 例文の単文の一覧

理により文脈処理を行った．文脈処理の結果としては，照応関係・一貫性関係・段落関係・単文の分類・フレーム・テーマ・フォーカス（後述）を出力する．図-4 に COGITO の文脈処理結果の意味表現の体系を示す．

文脈処理は，中間表現テキストを入力とし，要約ユニット間に関係をつけ，中間表現の概念を文書全体のネットワークにする．以下，図-5 に示すような新聞記事を例にして COGITO における文脈処理について述べる．図-6 に例文を単文に切った一覧を示す．本システムでは，トップレベルの単文の後に，埋め込まれた単文を depth-first の順に並べて処理している．

3.1 文脈処理の処理概要

COGITO の文脈処理は，名詞句解析・照応処理・一貫性処理・曖昧性解消・要約ユニット分類・テーマ管理・段落処理の順に実行する．以下，各処理について述べる．

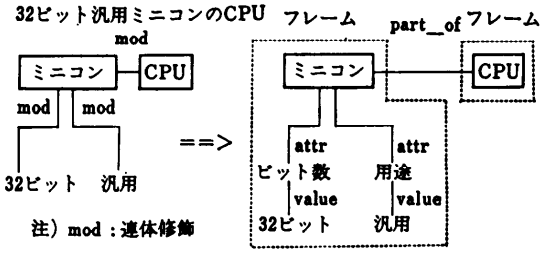
3.2 名詞句解析処理

照応処理での処理の効率化のために，中間表現名詞句に含まれる概念が対象-属性-属性値の組になるように概念を補い，同一名詞句中の対象同士を結びつけて

見出し：販売する

<格>	<存在する概念の意味素性>	<対象>
行為者格	φまたは組織	会社
	人	人
目的格	製品	製品
終点格	組織または人	ユーザ

図-7 用言概念辞書の例



注）mod：連体修飾

注）●□：対象
●part_of：全体-部分，attr：対象-属性，value：属性-属性値

図-8 フレーム化の例

フレームを生成する．フレーム化には 2. で述べた世界知識，及び，図-7 のような用言概念辞書を用いる．図-8 にフレーム化の例を示す．この処理において，概念が省略されている場合には，まず，格及び名詞句の中心語の意味素性をもとに用言概念辞書を用いて，名詞句の中心となる対象を制限し，次に，世界知識を用いて，省略されている各種概念を補足する．省略に対して補足できる概念の候補が複数個あるときは，世界知識において，総ての候補概念の上位にある概念を用いる．

3.3 照応処理

本システムでは概念同士の参照関係を照応関係，参照する側の概念を照応形，参照される側の概念を先行詞と呼ぶ．本システムでは，図-9 に示すような，Hirst[11] を基にした照応形の分類を用いた．本システムでは，名詞だけでなく，代名詞も概念に変換されているため，照応処理としては照応形の種類に関係なく，同じ処理になる．一方，照応関係が成り立つかどうかは，二つの概念だけから決まるわけではなく，概念を修飾している概念や概念を支配している用言の概念などを含めて考える必要がある．このため，照応はフレーム同士の関係として処理している．この場合，個々の概念同士の照応関係は，フレーム同士の照応関係から間接的に得ることができる．以下では，照応形・先行詞・照応関係という言葉をフレームに対して用いることにし，図-9 の照応形もフレーム化した形式で考えることにする．フレーム間の照応関係として

(i) 代名詞的 (pronominal)
 (a) 代名詞 (pronoun)
  単語や句を指す代名詞. ex. それ, これ
 (b) 代名詞的名詞 (Pronominal noun)
  単語や句を指す代名詞的名詞. ex. 同機, 同社
 (c) 別名 (epithets)
  あだ名などを表す名詞. ex. 近代音楽の父
 (d) 修辞的名詞 (surface count)*)
  修辞的に何かを指す名詞. ex. 両者, 前者, 後者
(ii) 動詞指示 (proactional)*)
  動作を指す単語. ex. そのこと, そうすること
  %前文指示 (prosentential) を含む.
(iii) 形容詞・関係指示 (proadjectinal/prorelative)*)
  属性や関係を指す単語. ex. そのような
(iv) 時間指示 (temporal)*)
  時間を指す単語. ex. そのとき
(v) 場所指示 (locative)*)
  場所を指す単語. ex. そこ
(vi) 省略 (ellipsis)
  行為者格・対象格の省略. ex. A の開発を進め, φ が φ を出荷する.
(vii) 限定照応形 (definite anaphora)
  名詞同士の関係. ex. その会社
*) 印はサポートしていない

図-9 照応形の分類[11]

表-2 照応関係の分類

照応関係名	定 義
equal	一致
is_a	上位—下位関係 (照応形が下位)
upper_of	上位—下位関係 (照応形が上位)
part_of	全体—部分関係 (照応形が部分)
include	全体—部分関係 (照応形が全体)

は表-2 に示すような関係を用いる. 本システムでは, 図-7 に示したように, 用言概念辞書に記述されている格のフレームについてのみ照応を考える. 照応は, すでに生成されているフレームとの照応関係をチェックし (現在は 50 文以内の単文, または, 直前の段落まで), それぞれの格のフレームごとに, 先行詞となる可能性のあるフレームの識別子と照応関係名の組をすべて生成する. 照応のチェックは, 以下の①〜③の順に行う.

① 対象同士が表-2 の照応関係をもつ.

② [スロットの無矛盾性]
一意の値しかもたない属性の値が異ならない.

③ [全体-部分フレームの無矛盾性]
part_of のリンクで結び付いたフレームについて, ①〜③を再帰的に適用する.

ただし, 本処理が終了した段階では書き換えは行わず, 後述する曖昧性解消処理で照応を一意に決定したうえでフレームやフレーム間のリンクの書き換えを行う.

3.4 一貫性処理

単文間の関係については, 一貫性[12]の考え方があるが, 要約システムでは, 節の初めに述べたように, 文書全体に適用できることを重視した. そのため, 本システムでは, 一貫性を広く解釈し, 単文の間に認められる関係を総称して一貫性関係と呼ぶことにした. できるだけ多くの観点から一貫性関係を定義することにより, 文書全体にわたって関係が得られる.

本システムの一貫性関係は, 以下に述べるような, 比較的浅い三つのグループの関係を定義した. Iグループは, 主に接続語 (接続詞・接続助詞・接続助詞的格助詞・イディオム) などを用いて単文同士の関係を決定する. この関係は, 隣接しない単文同士にも適用できる. IIグループは, 埋め込み文と主文の関係を埋め込み文の役割により分類した. たとえば, 連体修飾文について, 被修飾語が固有名詞の場合には, 連体修飾文は説明的で, 省略しても文の意味は変わりにくいが, 被修飾語が普通名詞の場合には, 連体修飾文は限定的で, 省略した場合, 文の意味が曖昧になってしまう. また, 中間表現により判定できる埋め込み関係として, 引用関係や同格関係などがある. IIIグループは, 照応に基づく関係でたとえばある単文のテーマ (単文の話題を表すフレーム. 3.7 で後述する) が前の単文のフレームを照応している場合に, 後の単文が前の単文の説明になっていると定義する. 一貫性関係のグループは, 独立である. 表-3, 表-4, 表-5 にそれぞれのグループの一貫性関係名の一欄を示す.

本システムでは, 着目している単文との間に一貫性関係をもつ可能性のあるすべての単文の番号と一貫性関係の組を生成する (現在は, 照応関係と同じ制限範囲).

3.5 曖昧性解消処理

一般に照応関係と一貫性関係は相互に依存関係があり, どちらかを先に決めることは難しい. たとえば, 照応処理により省略を補わないと一貫性の判定が決定できない場合があり, 逆に, 一貫性関係により省略が決定できる場合がある. さらに, 実体の異なる二つの

表-3 Iグループの一貫性関係

分類	個別の関係名		依存関係	関係の内容
接続語による関係	仮定		--〉,〈--	一方の単文が仮定になっている
	原因・理由		--〉,〈--	一方の単文が原因・理由になっている
	連続		〈--〉	二つの単文が連続した事象になっている
	並列		〈--〉	同趣の事柄を並べあげる
	補足		〈--	前文を補足する
	時	前	--〉,〈--	一方の単文が他方の前に起こる
		同時	--〉,〈--	一方の単文が他方と同時に起きる
		後	--〉,〈--	一方の単文が他方の後に起きる
		継続1	--〉,〈--	一方の単文が継続中に他方が起こり続ける
		継続2	--〉,〈--	一方の単文が継続中に他方が起きる
default	連続		〈--〉	二つの単文間に他の接続関係がない

表-4 IIグループの一貫性関係

分類	個別の関係名	依存関係	関係の内容
連体修飾及び引用	接続	〈--	前文と主文を接続する
	説明	--〉	被修飾語についての説明
	限定	--〉	被修飾語を限定する
	内容	--〉	被修飾語の内容になっている
	引用	--〉	引用になっている

表-5 IIIグループの一貫性関係

分類	個別の関係名	依存関係	関係の内容
照応	説明（照応関係）	〈--	単文1の単語と単文2のテーマに照応関係がある
	内容	--〉,〈--	一方の単文中の単語の内容が，他方の単文になっている

照応形が同じフレームを先行詞の候補とする場合がある．このため，本システムでは，照応処理と一貫性処理については，すべての候補を生成し，曖昧性解消処理により，照応関係と一貫性関係，あるいは，照応関係同士に矛盾が起きないようにして，これらの関係を一意に決定する方式を採用した．現在は，新聞記事を対象とした簡単なルールを用いている．以下に，ルールの例を示す．

① 照応関係の曖昧性解消ルール

以下のルールを上から順に適用する．

（a） 過去にテーマ又はフォーカス（3.7 参照）になったフレームを優先する．

（b） 全体-部分，上位-下位では，全体-部分を優先する．

（c） もっとも近い関係を優先する．

（d） 二つの照応関係が矛盾するときは，文の先頭に近いほうの照応形を優先する．

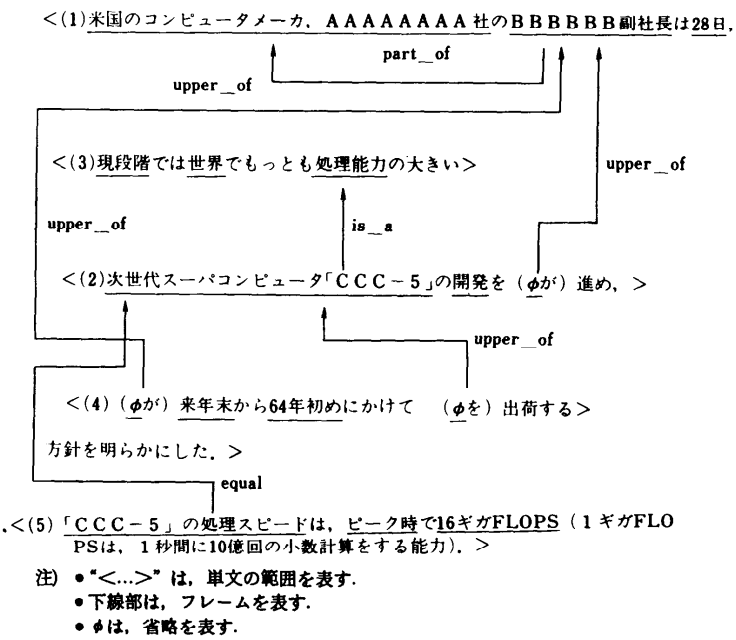
② 一貫性関係の曖昧性解消ルール

- Iグループは，もっとも近い関係を選ぶ．
- IIグループは，曖昧性なし．
- IIIグループは，照応により定義しているため，照応の結果に依存する．

図-10 に例文の照応処理結果の一部を示す．また，図-11 に例文の一貫性処理結果を示す．照応と一貫性の相互依存関係は，たとえば，並列文と思われる二つの文において，後の文の主語が省略されている場合に，主語が等しいかどうかのチェックから並列文と判断するか，並列文かどうかのチェックにより，主語を等しいと判断するかといった問題があり，今後の検討を要する．

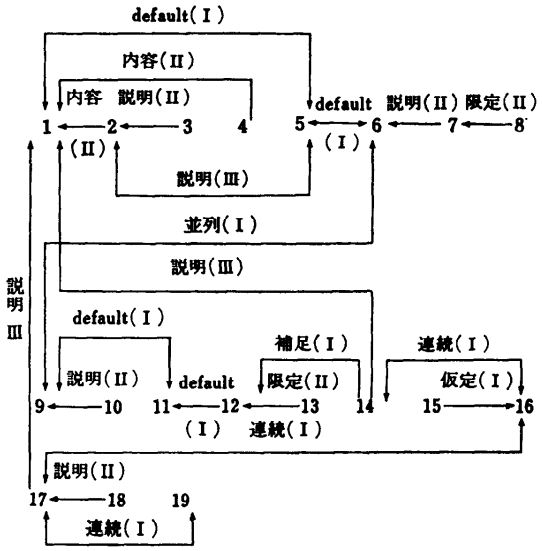
3.6 要約ユニット分類ルーチン

単文を事実・意見に分類する．



注）
- "<...>" は，単文の範囲を表す．
- 下線部は，フレームを表す．
- φは，省略を表す．

図-10 例文の照応処理結果の一部



注) ●番号は単文番号
●リンクの関係名は表-2, 3, 4 参照
●括弧内は一貫性関係のグループ名 (3.4 参照)

図-11 例文の一貫性処理結果

3.7 テーマ管理ルーチン

一般的に，文には，前提となる部分と新たな事実が述べられる部分がある．COGITO では，前提となる部分の中心的なフレームをテーマ，新たな事実が述べられる部分の中心的なフレームをフォーカスと定義している．テーマは次のようなルールで決定している．

① 提題の格（日本語では，「は」のつく格）になっているフレーム

② ①がない場合，前の単文と照応関係をもつフレームのうち単文の先頭に近いフレーム

③ ①，②がない場合，単文の先頭のフレーム

④ ①，②，③がない場合，直前の単文と同じ

また，フォーカスは，次のようなルールで決定している．

① テーマが照応関係をもった場合に，先行詞をその先行詞を含む単文のフォーカスにする

3.8 段落処理ルーチン

照応関係が一貫性関係に利用できるのと同様に，段落にまたがる一貫性関係は，単に単文同士の関係としてだけではなく，段落同士の関係に利用できる．本システムでは，段落にまたがる一貫性関係のうち，後の段落で最初に出現する一貫性関係を二つの単文の属する段落間の関係にする．

以上のような文脈処理により，照応関係・一貫性関係・段落関係・テーマ・フォーカス・単文の分類が生成される．

4. 重要性の評価方式と評価結果の出力

4.1 重要性評価

重要性評価では，ユニットの重要性評価を行う[13]．重要性評価には，さまざまな方法が提案されている．本システムでは，1. で設定した重要性の定義を基にして，文脈処理の結果及び表層的な情報を用いて，具体的な重要性評価の方法を検討した．以下に，重要性評価のルールの例を示し，ルールの内容及び設定理由を 1. で述べた重要性条件と関連づけて説明する．現在，ルールはすべて独立に適用され，各ルールの評価結果 $E_1, E_2, \cdots$ の重みつきの総和が各ユニットの重要性評価値 E となる．現在，ルールの重みは，すべて対等にしている．以下で，評価値といった場合には，重要性評価値を表す．なお，ここでは，重要性評価の対象となる要約ユニットを単文に限定している．

（1） 文脈処理結果による重要性評価

ルール 1 ：一貫性による重要性評価

一貫性関係により，単文の依存関係が得られる．それぞれの単文に依存している単文の数をカウントすることにより，隣接する単文との相対的な重要性が計算できる．多くの単文から依存されている単文は，省略すると文書の意味が曖昧になる可能性が高いため，重要性の曖昧回避条件により重要と考えられる．さらに，重要な単文が依存している単文は重要であることが予想されるため，依存されている単文の重要性を一定の割合で加えることにより，間接的に依存している単文の重要性の伝播を行う[8]．

単文の評価値 E_1 を次のようにして決める．

$$E_1 = \sum_{n=1}^{\infty} P_n \times N_n$$

N_n：評価する単文に，一貫性関係による依存関係で n 回のリンクをたどって間接的に依存する単文の総数

P_n：リンクの伝播係数（n の増加にともない減少する）$P_n = (1/2)^{n-1}$ としている．

ルール 2 ：埋め込み文による重要性評価

埋め込み文は，それが修飾している語句の説明であることが多く，その場合は除いても文の意味にあまり影響がない．ただし，限定の意味をもつ連体修飾文（たとえば，「A社の発表したパソコン」，というような単文）はそれを除いてしまうと「パソコン」の示す

対象が総称的になり，「パソコン」の意味する範囲が変わってしまうため重要性が高い（曖昧回避条件）．このため，単文の評価値 E_2 として，限定の埋め込み文は0，その他の単文は -1 とする．

ルール3：文内の概念から構成されるフレームによる重要性評価

他の文とつながりがなくても情報を多く含む文は重要と考えられる（復元困難条件）．このため，単文の評価値 E_3 として，文内の概念から生成されたフレームのスロット数を評価値とする．

ルール4：単文の分類による重要性評価

単文の分類は，一文で決定できるが，事実か意見かにより文書中で異なる役割が与えられる点では，文脈的である（曖昧回避条件）．この評価は文書の種類に依存する．たとえば，評価値 E_4 は，報告文については，事実を述べている単文は1，その他の単文は0とし，論説文については意見を述べている単文は1，その他の単文は0とする．

（2） 表層情報による重要性評価

一貫性関係と照応関係だけでは，重要性評価の対象としては十分ではなく，また，解析時の誤差も存在するため，表層的な重要性評価を併用する．

ルール5：分野ごとの重要語による重要性評価

対象や属性の中でも重要性の高いものとそうでないものがある（復元困難条件）．あらかじめ分野ごとに重要語を決めておき，単文の評価値 E_5 として，重要語の概念を含む単文に1，その他の単文に0を与える．製品紹介記事では，「価格」，「特徴」，「目的」，「販売」，「開発」，といった語は，重要である．

ルール6：修辞的な語句による重要性評価

「要するに」，「結局」などは，作者の意図が反映されているため（意図伝達条件），単文の評価値 E_6 として，このような語を含む単文は1とし，逆に，「たとえば」，「ただし」などを含む単文に -1，その他の単文に0を与える．

ルール7：強調表現

強調表現は，作者の意図が反映されているため（意図伝達条件），このような語を含む単文は重要である単文の評価値 E_7 として，強調表現の数を評価値とする．強調表現の例は以下のようなものである．強調表現としては，「非常に」，「まったく」，「〜さえ」などがある．

（3） ユニットの重要性の計算方法

重要性の計算は，ユニットごとに，重要性評価ルール1〜7によって決定された重要性に重みをつけて加算する．すなわち，あるユニットの重要性 E は，

$$E=\sum_{i=1}^{7} W_i \times E_i$$

で求められる．ここで，W_i はルール i の評価値に対する重みで，ルールがユニットに対して適用できないときは，0とする．E_i はルール i による評価値である．重みづけは変更することができ，それにより柔軟な重要性評価が可能となる．

4.2 重要文の抜粋・表示

重要性評価値の大きい単文を抜粋し，分かりやすく表示する．一般に抜粋部分を羅列するだけでは誤解を生じる場合があるため，抜粋部分の再構成が必要である．ここでは，要約文を抜粋部分の表示方法と広義に解釈すると，要約文の形態としては，以下のようなものが考えられる．

① 原文を表示して抜粋部分にマークづけをする．

② 抜粋部分を一部修正し，抜粋部分のみを表示する．

③ 抜粋した語句を表形式で表示する．

④ 抜粋部分を内容とする文章を生成する．必要に応じて語句の置き換え，生成順序の変更，代名詞化，必須格の補充などを行う．

これらは，用途によっても異なり，必ずしも優劣はつけられない．さらに，抜粋された部分においても重要性の高い部分とそうでない部分があり，表示の際に表現方法を考慮する必要があると考えられる．COGITOでは，方法①，②で表示している．

4.3 重要性評価例

例文に対して重要性評価を行い，単文を抜粋した結果を図-12に示す．また，表-6は，重要性評価の結果である．図-12では，10点以上の評価値の単文を抜粋した．本システムは，支援システムであるため，何点以上の単文を抜粋するかは，縮小率（例では，字数の縮小率が40%）や全体のバランスをみながら利用者が指定できるようにしている．ルール5は，「開発」，「出荷」，「販売」，「処理スピード」に適用され，また，ルール6は，「ても」，「最も」に適用されている．なお，図-12の抜粋部分だけを表示した場合に意味が不十分になる部分（たとえば抜粋結果3行目の「方針」

米国のコンピュータメーカ，AAAAAAAA社のB！
BBBBB副社長は28日，現段階では世界で最も処理能！
力の大きい次世代スーパコンピュータ「CCC—5」の！
開発を進め，来年末から 64 年初めにかけて出荷する！
方針を明らかにした．
「CCC—5」の処理スピードは，ピーク時で16ギガ！
FLOPS（1ギガ FLOPS は，1秒間に 10 億回の小数！
計算をする能力）．現在，世界で販売されているスー！
パコンピュータで最も処理スピードの速い米国DDD！
DDDDDDD 社の EEE—10E モデル（6.9 ギガ！
FLOPS）の2倍以上となる．また，AAAAAAAA社！
製品で最も処理スピードの速い「CCC—4」（1.95 ギ！
ガ FLOPS）の約8倍．記憶容量も 5，6 倍になるとい！
う．わが国で稼働中のスーパコンピュータは現在，約！
100 台といわれ，うちAAAAAAAA社の設置台数は！
7台，受注済みを含めても 11 台．BBBBBB副社長！
は，米商務省の主催で 27 日から東京，大阪で商談セミ！
ナを行っている米国スーパコンピュータ貿易使節団の一！
員として来日，官公庁を中心に売り込みを続けている．！

図-12 抜粋の例（評価値10以上の単文）

表-6 重要性評価結果の例

単文番号	重要性評価値							
	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E
1	11.50	0	4	1	0	0	0	16.50
2	5.00	0	2	1	1	0	0	9.00
3	0	−1	2	1	0	0	0	2.00
4	0	0	2	1	1	0	0	4.00
5	7.50	0	3	1	1	0	0	12.50
6	6.75	0	3	1	0	0	0	10.75
7	1.00	−1	1	1	0	0	1	3.00
8	0	0	1	1	1	0	0	3.00
9	6.00	0	2	1	0	0	1	10.00
10	0	−1	2	1	1	0	1	4.00
11	5.00	0	1	1	0	0	0	7.00
12	5.50	0	1	1	0	0	0	7.50
13	0	0	1	1	0	0	0	2.00
14	2.25	0	2	1	0	0	0	5.25
15	0	0	1	1	0	0	1	3.00
16	4.25	0	1	1	0	0	0	6.25
17	4.25	0	3	1	0	0	0	8.25
18	0	−1	3	1	0	0	0	3.00
19	2.75	0	0	1	0	0	0	3.75

など）があるが，これは，Ⅱグループの一貫性関係において「内容」や「限定」関係の場合には，重要性評価値にかかわりなく，埋め込み文を抜粋に加えることで対応できる．

要約結果を評価するためには基準が必要であるが，要約の定義自身が明確でなく，要約の用途により，異なった基準が考えられる．ここでは，人間が重要とした部分と，システムによる結果を比較した．図-13 に人間と要約支援システムによる重要性評価結果の比較を示す．人間による重要性評価は，一人が要約文に選んだ場合に1点とし，4人の評価を合計したものである（単文の一部を抜粋した場合は0.5点としている）．この評価例では，一貫性による重要性評価値が大きくなっている．一般に，一貫性による評価では，文書の全体に評価値が平均的に分布し，埋め込み文の評価値は低くなる．本例では全員が単文4を重要としているが，システム側の評価は低い．これは人間は出荷時期が重要と判断したためであり，このギャップは重みづけの変更で調整可能である．

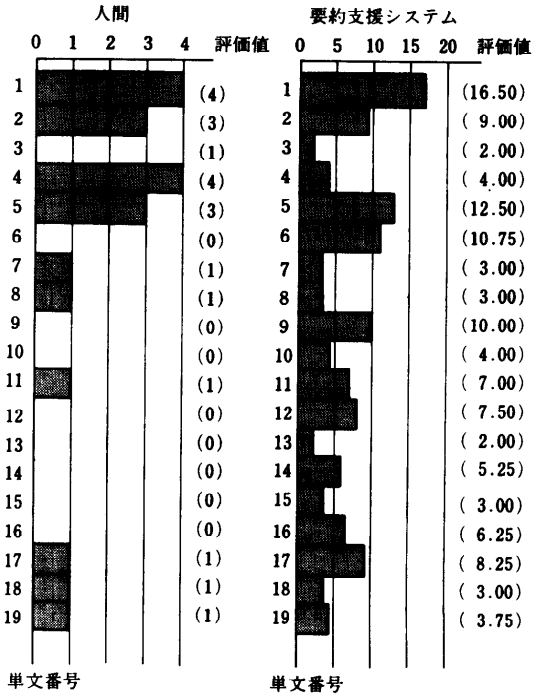


図-13 人間と要約支援システムによる重要性評価結果の比較

5. おわりに

本システムは主に製品の紹介記事を対象として開発したものである．前提条件として，入力文は現在の機械翻訳で使用されている中間表現レベルまで解析されたものと仮定している．それに対して，ここで述べた文脈的関係を用いた要約アルゴリズムは，用言については用言概念辞書を，体言については世界知識を用いているが，これらの知識は，電子化辞書[14]の開発によ

り実現可能性が高く，要約モデルの一つの解を与えるものと考えている．

他方，論説文の要約過程は，複雑であり，多くの知識や高度の推論が必要となるが，これらはいまだ十分にアルゴリズム化されていない．このような要約を実現するためには，背景にある省略された知識の補充や，より深い知識表現の利用など，文書の内容に重点をおいた文書解析，及び，主題の解析や論旨の把握などにより理解を深めていくような重要性評価が必要となり，これらは，状況意味論，談話理解など今後の言語理論の発展に待たなければならないと考える．

なお，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行われた．

謝辞 本研究を支援していただいた ICOT 渕所長，内田室長，吉岡室長代理に感謝いたします．

参 考 文 献

1) 北，小松，安原：要約システム COGITO，情報処理学会研究報告 86-NL-57 (1986).
2) 小松，加藤，安原，椎野：要約システム COGITO—文書の構造解析—，情報処理学会研究報告 87-NL-64 (1987).
3) Luhn, H. P.: The Automatic Creation of Literature Abstracts, IBM Journal, Vol. 2, No. 4, pp. 159-165 (1958).
4) DeJong, G.: Prediction and Substanination, A New Approach to Natural Language Processing, Cognitive Science 3, pp. 251-273 (1979).
5) 猪瀬，斎藤，堀：シナリオを用いる論文抄録理解・作成援助システム，情報処理学会論文誌 Vol. 24 No. 1 pp. 22-29 (1983).
6) Rumelhart, D. E.: Notes on a Schema for Stories, in Bobrow & Collins, eds., Representation and Understanding, Academic Press (1975).
7) 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).
8) Hasida, K., Isizaki, S. and Isahara, H.: An Approach to Abstract Generation, Bul. Electrotech. Lab., Vol. 52, No. 4, pp. 551-564 (1988).
9) Lenhert W. G.: Plot Units and Narrative Summarization, Cognitive Science, Vol. 5, No. 4 (1981).
10) 田村：要約過程の形式化と実現について，人工知能学会誌，Vol. 4, No. 2, pp. 198-206 (1989).
11) Hirst, G.: Anaphora in Natural Language Understanding A Survey, Lecture Notes in Computer Science 119 (1981).
12) Hobbs, J. R.: On the Coherence and Structure of Discourse, Center for the Study of Language and Information, Report No. CSLI-85-37 (1985).
13) Fum, D., Guida, G. and Tasso, C.: Evaluating Importance: A Step towards Text Summarization, IJCAI-85, pp. 840-844 (1985).
14) 日本電子化辞書研究所：概念辞書（第1版）/単語辞書（第2版），Technical Report TR-006/TR-007 (1988).

（平成元年6月24日受付）