

要約システムCOGITO

安原 宏、小松 英二、日比 孝、加藤 安彦

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

沖電気工業株式会社総合システム研究所

株式会社日本電子化辞書研究所（本解説は沖電気工業
株式会社に在職中の成果に基づいてなされたものであ
る）

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・テキスト内の位置等を用いて、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フレームや文法で記述しておき、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^{4)・5)・6)}。ボトムアップ的な要約では、文書を談話構造等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)・8)・9)・10)}。以上のように、要約には様々な方法が提案されているが、これらをまとめると、図-1に示すようになる。人間による要約も、詳細な理解を伴う要約と、斜め読みで文書の大筋を理解するような2種類の要約があると考えられるが、COGITOでは、これらを併用したシステムを目標にしている。

要約で第一に問題になるのは、重要性の定義である。Luhn³⁾は文書の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、文書の特徴づける単語、文、パラグラフの存在が考えられる。COGITOでは、単語、単語、文、パラグラフを総称して、重要性評価の単位になるという意味で、「要約ユニット」或いは単に「ユニット」と呼び、文書の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、机上の検討により、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）等で多く結び付いているユニット。復元困難条件は、意外性や特異性のある事象、数値、固有名詞等のユニット。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（例えば、新聞記事では5W1H等）。指示条件は、一般的に読者が関心を持つユニットや、個人的に興味を持っているユニット。上記の4つの条件は複合的な条件であり、1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3つの条件の重要性評価も行っている。

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-2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-3に示すように、格フレームを概念と概念の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、パラグラフ番号、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)を表わす。

(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・パラグラフ関係(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、事実か意見かについての重要性評価を行うために単文の分類を決定する。

(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、ルール毎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付きの総和として、各ユニットの重要性評価値を算出する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程重要であり、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

世界知識は図-4に示すように、概念を階層構造にしたものである。概念はフレームの見出しとなる対象、スロットとなる属性、スロット値になる属性値に分類され、概念同士は上位一下位、全体一部分、対象一属性、属性一属性値といった関係で結ばれている。辞書は、単語の概念や文法情報を記入した解析辞書、及び、省略等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付けられる。

3. 文脈処理方式

文書中のユニットの重要度は、文書全体から相対的に判断されるものであり、文を超えた関係を解析することが必要である。

質の良い要約を行うためには、文書を深く解析することが必要であるが、同時に、要約システムの目的の一つは、予測のつかない多くの文書の概要を知ることであり、このためには大量の文書に対応できる汎用性のある解析手法が要求される。一般に文書における文は対話に比べ複雑な構造をしており、複雑な知識を適用することが難しい。また、内容的にも事前に範囲を制限することが難しい。このため、COGITOでは、汎用性を満たすため、主に言語的な情報の解析を行う手法をとることにした。

本システムは、中間表現テキストを入力とし、ユニット相互の関係を認識するための文脈処理を行い、要約ユニット間に関係をつけて、中間表現の概念をネットワーク化する。図-5にCOGITOにおける意味表現の体系を示す。要約ユニット間の関係としては、中間表現の格関係に加えて、後述するような照応関係・一貫性関係・パラグラフ関係を用いる。図-6に以下の各節で解析の例として用いる新聞記事の原文を示す。また、図-7に例文を単文に切った一覧を示す。単文は文脈処理の単位である。以下に、本システムにおける文脈的意味表現及び文脈処理方式について述べる。

3.1 文脈処理の処理概要

図-8にCOGITOの文脈処理の処理フローを示す。COGITOの文脈処理は、要約ユニット分類・名詞句解析・照応処理・一貫性処理・曖昧性解消・テーマ管理・パラグラフ処理からなる。以下、各処理について述べる。

3.2 名詞句解析処理

本システムでは、照応処理¹¹⁾を行うが、一般に照応は単語同士だけのチェックでは不十分で、単語の修飾語や単語の属する単文の用言等も含めて処理する必要がある。COGITOでは、これらの情報を考慮できるようにするため、照応に先立って、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って、フレームを生成する。また、同一名詞句中の対象同士も結び付ける。フレーム化には2節で述べた世界知識、及び、図-9のような用言概念辞書を用いる。図-10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。ここで生成したフレームは、照応処理における書き換えにより、さらに詳細化される可能性がある。

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と定

義した。また、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-11に示すような、Hirst¹⁾を基にした照応形の分類を用いた。本システムでは、代名詞等は単語が表わす範囲に最も近い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。本システムでは、フレーム同士の関係として照応処理を行っており、個々の概念同士の関係は、フレーム同士の関係から間接的に得られる。このため、以下では、照応形・先行詞・照応関係という言葉はフレームに対して用いることにする。照応関係としては表-1に示すような関係を用いる。本ルーチンでは、着目している単文のフレームと既に生成されているフレームとの照応関係をチェックする（現在は50文以内の単文、または、直前のパラグラフまで）。単文中の必須格のフレームに対し、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総て生成する。照応のチェックは、フレーム同士のユニフィケーションにより、以下のi)～iii)の順に行う。

i) 対象同士が表-1の照応関係を持つ。

ii) [スロットの無矛盾性]

一意の値しか持たない属性の値が異ならない。

iii) [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、

i)～iii)を再帰的に適用する

ただし、本処理が終了した段階ではユニフィケーション

ョンによる書き換えは行わず、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、重要性評価が正しく行える
ことが目的であり、必ずしも深い内容を反映した関係
は必要でない。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。

一貫性関係は、節の初めに述べたように、汎用性を
持たせるため、以下に述べるような、比較的浅い3つ
のグループの関係をを用いる。1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）等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。2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。例えば、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。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
係等がある。3グループは、照応に基づく関係で、例
えばある単文のテーマ（単文の話題を表わすフレーム。
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-2、表-3、表-4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持つ可能性のある総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

一般に照応の先行詞や一貫性関係を持つ単文を一意に決定するためには、照応関係と一貫性関係の相互関係、文書の種類、言語外の情報等が必要である。例えば、新製品の発表記事では、文を通して新製品が話題であることが前提になっており、「価格」という単語が文書のどこに現われても、新製品と参照関係を持つ可能性が高い。

本システムでは、一貫性処理で生成した照応関係及び一貫性関係の候補の中から最も優先度の高い組み合わせを選び、照応関係のあるフレームの書き換え・リンクづけ及び単文同士のリンクづけを行う。現在は、照応関係と一貫性関係が相互に依存する場合は扱っておらず、以下のような簡単なルールを用いている。

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。

- ・過去にテーマ又はフォーカス（3.7 節参照）になったことのあるフレームを優先する。

- ・全体－部分、上位－下位では、全体－部分を優先する。
- ・最も近い関係を優先する。
- ・照応同士が矛盾するときは、文の先頭に近いほうの照応を優先する。

②一貫性関係の曖昧性解消ルール

- ・グループ1は、最も近い関係を選ぶ。
- ・グループ2は、曖昧性なし。
- ・グループ3は、照応の曖昧性解消結果に従う。

図-12に例文の照応処理結果の一部を示す。また、図-13に例文の一貫性処理結果を示す。

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。

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
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持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 テーマから照応を受けているフレーム

3.8 パラグラフ処理ルーチン

照応関係が一貫性関係に利用できるのと同様に、パラグラフにまたがる一貫性関係は、単に単文同士の関係ではなく、パラグラフ同士の関係を表わしている場合がある。本システムでは、パラグラフにまたがる一貫性関係のうち、後のパラグラフで最初に出現する一貫性関係を2つの単文の属するパラグラフ間の関係にする。

以上のような文脈処理により、照応関係・一貫性関係・パラグラフ関係・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、文書解析の結果を用い、ユニットの重要性評価を行う¹³⁾。重要性評価には、様々な方法が提案されている。本システムでは、1. で述べたような重要性の定義に基づき、文書解析における文脈処理の結果と表層的な情報を併用して重要性評価を行っている。以下に、重要性評価のルールの例を示し、ルール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。ルールは、総て独立に適用され、各ルールの評価結果 $E_1, E_2, \dots$ の重み付きの総和が各ユニットの重要性評価値 E となる。以下で、評価値といった場合には、重要性評価値を表わす。なお、ここでは、重要性評価の対象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¹⁴⁾。

単文の評価値 E_1 を次のようにして決める。

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

↑
1/2ⁿ

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 の増加に伴い減少する)

$$P_n = (1/2)^{n-1} \text{としている。}$$

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、限定の意味を持つ連体修飾句（たとえば、「A社の発表したパソコン」、と言うような単文）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、そのたの単文は-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

他の文との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、文脈的であり（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E4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）。予め分野毎に重要語を決めておき、単文の評価値E₅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」、「価格」といった語は、重要である。

ルール6：修辭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E₆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」等を含む単文に-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_u 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」等がある。

(3) ユニットの重要性の計算方法

重要性の計算は、ユニット毎に、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、ユニット U の重要性 E_u は、

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

$\downarrow$ 74ル
 $\uparrow$ 74ル

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、ユニット U 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 要約文生成

要約文生成は重要性評価により抜粋された部分を分かりやすく表示する処理である。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、必須格の補充等を行う。

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、方式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、10点以上の評価点をとった単文を抜粋した結果を図-14に示す。また、表-5は、重要性評価の結果である。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。

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-15に人間と要約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。

5. おわりに

機械翻訳が解析型の問題であるのに対して、要約は設計型の問題とみることができる。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り、それに対して、ここで述べた文脈的關係を用いた要約アルゴリズムは、用言の用言概念辞書と体言の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1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、省略された背景にある事項の補足や、より深い知識表現の利用など、文書の意味的な構成に重点をおいた文書解析、及び、主題の解析や論旨の把握等により理解を深めていくような重要性評価が必要になり、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]

本研究を支援して頂いたICOT所長、内田室長、吉岡室長代理に感謝致します。

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 Substantiation,
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., 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 pp198-206 (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 Carlo:

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)

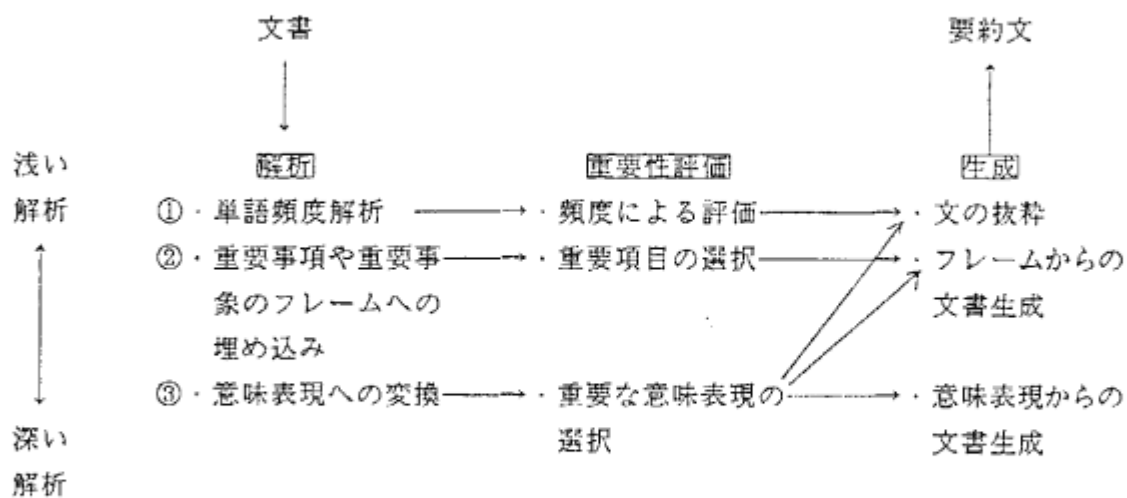


図-1 要約の方法

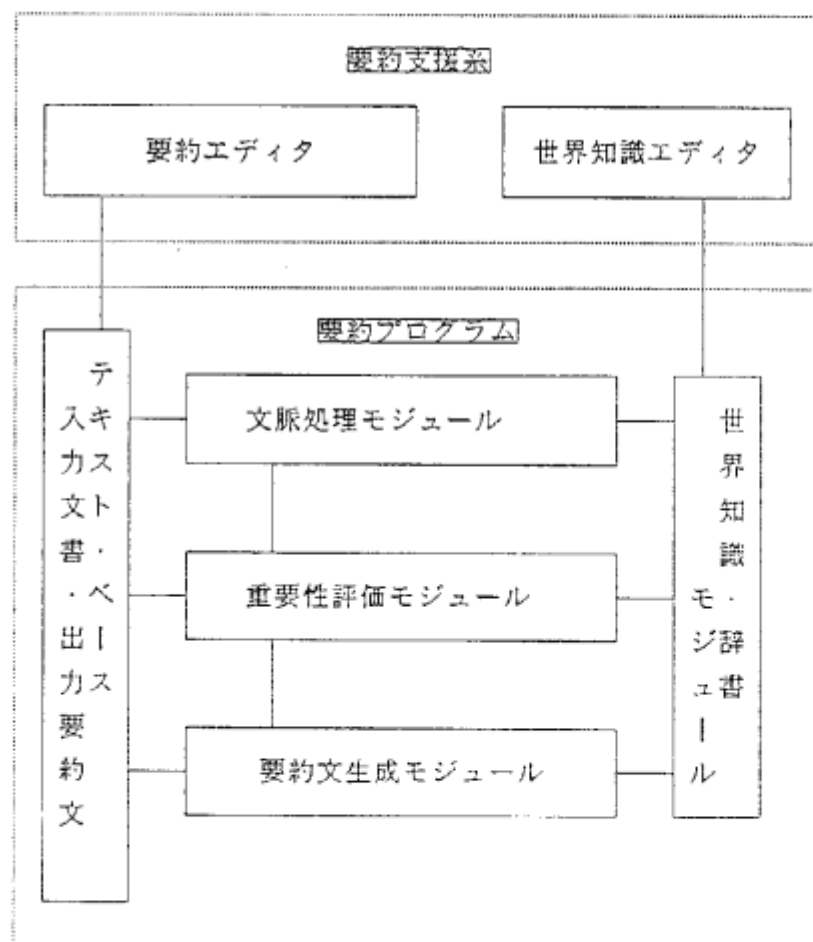


図-2 COGIT0の概略構成図

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],  
        [パラグラフ番号, 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,  
        [パラグラフ番号, 文番号, 単文番号, 被修飾語の単語番号]).
```

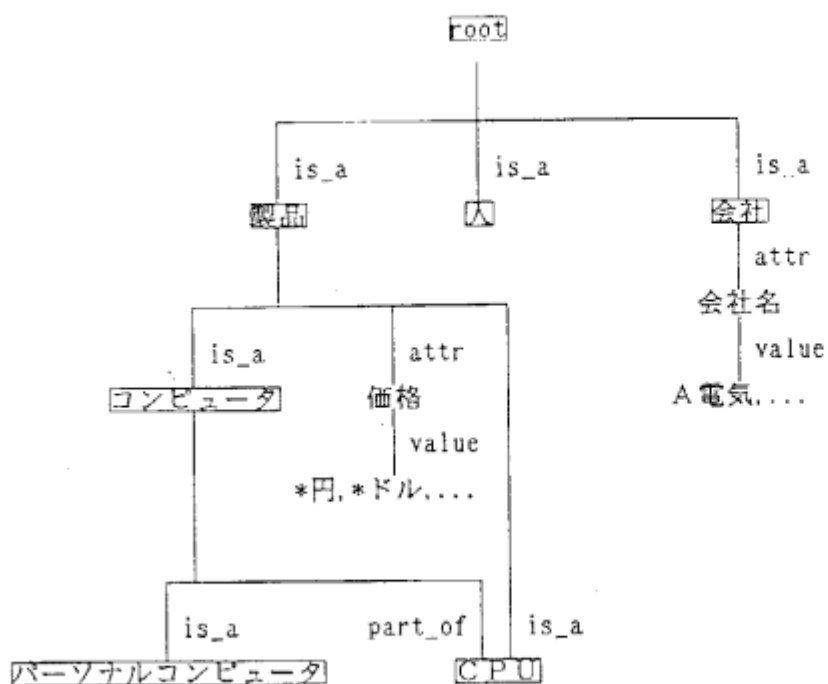
例 :

「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

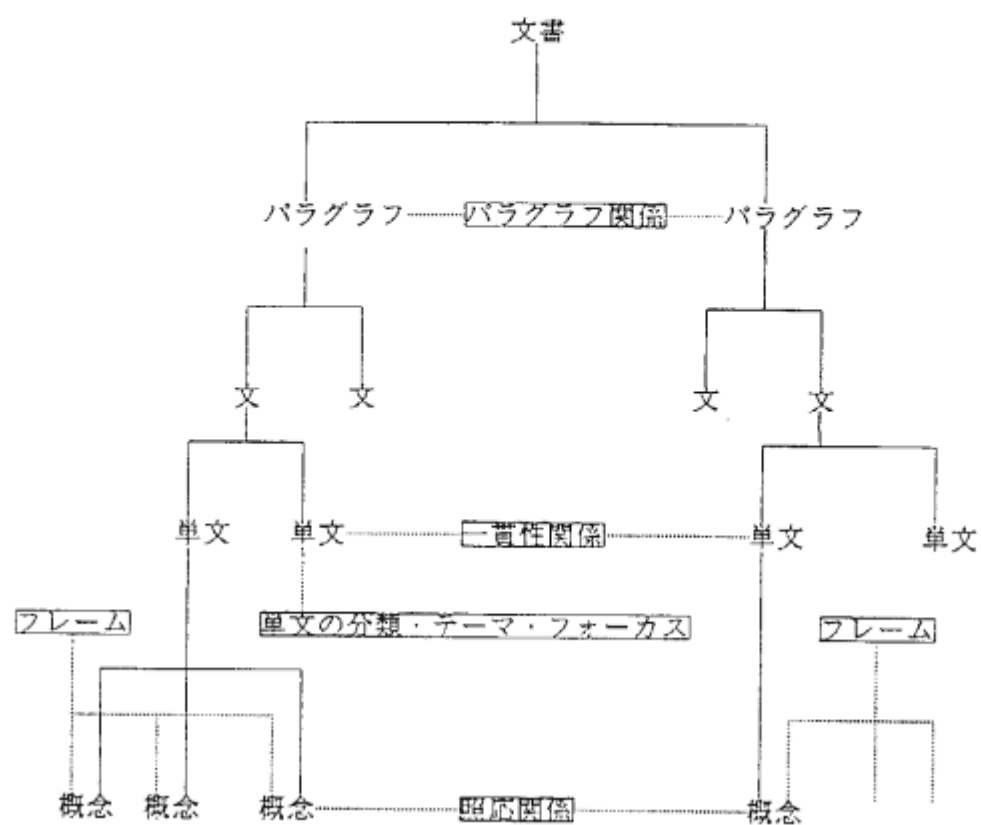
```
meaning([発表する, 6], agent, [A電気, 1], [1, 1, 1, 1]),  
meaning([発表する, 6], time, [[nil, 18, 日], 2], [1, 1, 1, 2]),  
meaning([RAM, 4], [mod, def], [[nil, 1, メガビット], 3], [1, 1, 2, 3]),  
  
meaning([生産する, 5], obj, [RAM, 4], [1, 1, 2, 4]),  
meaning([生産する, 5], tense, present, [1, 1, 2, 5]),  
  
meaning([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-3 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わし、
 その他のノードは属性・属性値を表わす
 ・ is_a: 上位-下位、part_of: 全体-部分、
 attr: 対象-属性、value: 属性-属性値

図-4 世界知識の例



注) □のないノード：中間表現の構造構造
 □付ノード：文脈的な意味表現

図-5 意味表現の体系

処理能力最大のスーパー電算機 88 年末にも発売 A A A A A A A A 社

米国のコンピューターメーカー、A A A A A A A A 社の B B B B B B 副社長は 28 日、現段階では世界最も処理能力の大きい次世代スーパーコンピューター「C C C - 5」の開発を進め、来年末から 64 年初にかけて出荷する方針を明らかにした。

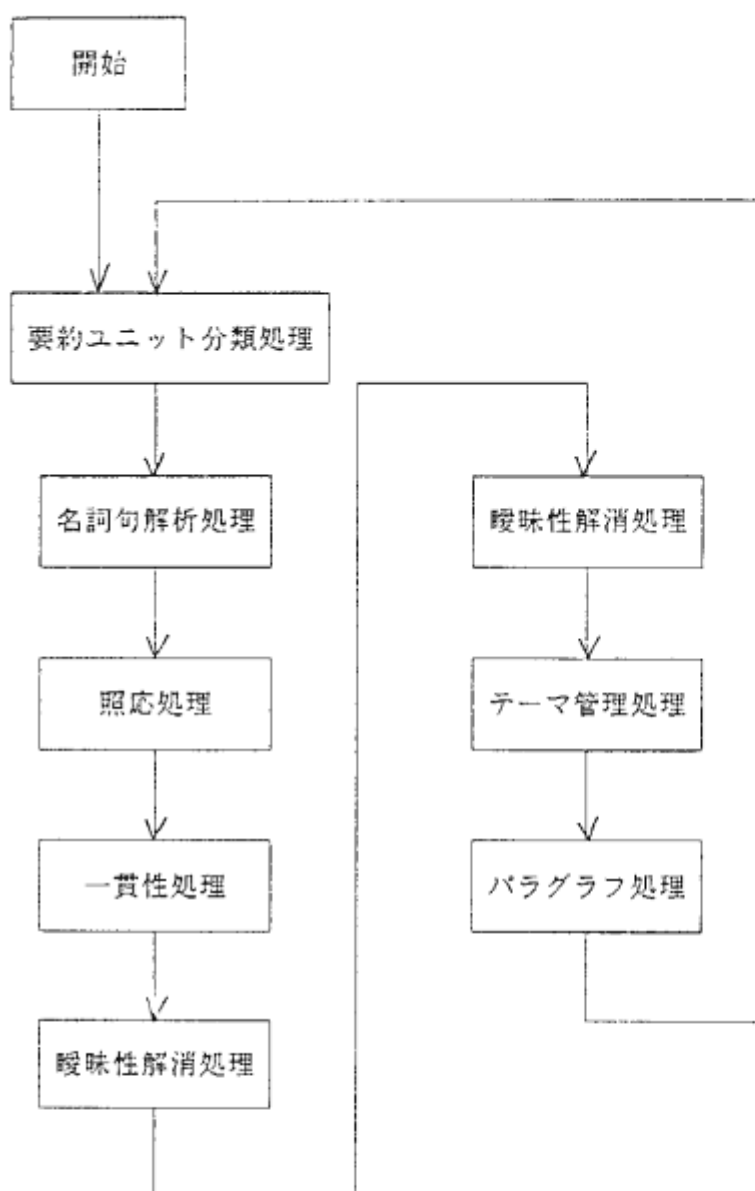
「C C C - 5」の処理スピードは、ピーク時で 16 ギガ F L O P S（1 ギガ F L O P S は、1 秒間に 1 億回の小数計算をする能力）。現在、世界で販売されているスーパーコンピュータで最も処理スピードの速い米国 D D D D D D D D D D 社の E E E - 10 E モデル（6.9 ギガ F L O P S）の 2 倍以上となる。また A A A A A A A A 社製品で最も処理スピードの速い「C C C - 4」（1.95 ギガ F L O P S）の約 8 倍、記憶容量も 5、6 倍になるという。わが国で稼働中のスーパーコンピューターは現在、約 100 台といわれ、うち A A A A A A A A 社の設置台数は 7 台、受注済みを含めても 11 台。B B B B B B 副社長は、米商務省の主催で 27 日から東京、大阪で商談セミナーを行っている米国スーパーコンピューター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-6 例文の原文

1. 米国のコンピューターメーカー、AAAAAAAAA社のBBBBBB副社長は28日、方針を明らかにした。
 2. 次世代スーパーコンピューター「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DDDDDDDDDD社製のEEE-10Eモデル（6.9ギガFLOPS）の2倍以上となる。
 7. スーパーコンピューター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. AAAAAAAAAA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. スーパーコンピューターは現在、約100台といわれ、
 13. わが国で稼働中の
 14. うちAAAAAAAAA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、米国スーパーコンピューター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、大阪で商談セミナーを行っている
 19. 官公庁を中心に売り込みを続けている。

注）単文は、一文内でトップの単文からdepth-firstで番号付けしている。

図-7 例文の単文の一覧



注) ← : 処理の流れ

図-8 文脈処理の概略処理フロー

1

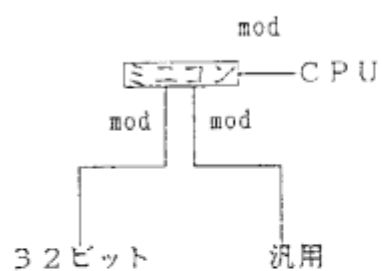
図-9 用言概念辞書の例

図-9 用言概念辞書の例

図-9 用言概念辞書の例

図-9 用言概念辞書の例

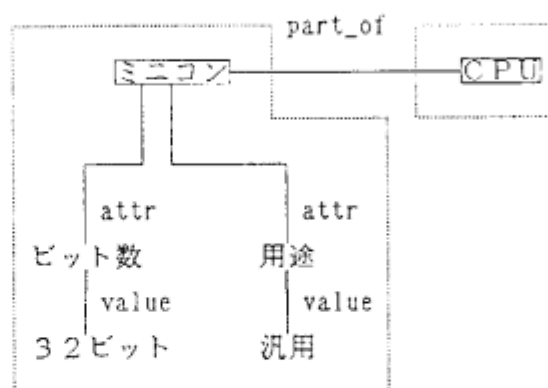
32ビット汎用ミニコンのCPU



==>

フレーム

フレーム



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-10 フレーム化の例

- (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)

あだ名等を表わす名詞。ex. 近代音楽の父
 - (d) 修辭的名詞 (surface count) *)

修辭的名に何かを指す名詞。ex. 兩者、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・關係指示 (proadjectival/prorelative) *)

属性や關係を指す単語。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)

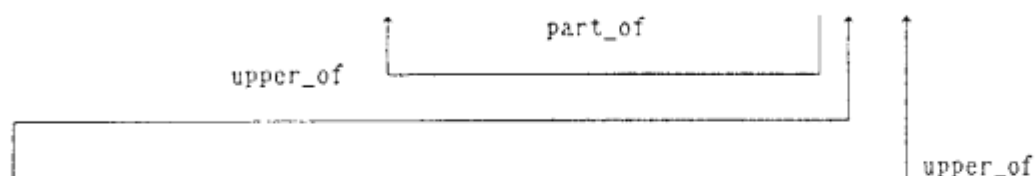
行為者格・対象格の省略。ex. Aの開発を進め、φがφを出荷する。
- (vii) 限定照応形 (definite anaphora)

名詞同士の關係。ex. その会社

*) 印はサポートしていない

図-11 照応形の分類¹¹⁾

<(1) 米国のコンピュータメーカー、AAAAA社のBBBBB副社長は28日、



<(3) 現段階では世界で最も処理能力の大きい>

upper_of

is_a

<(2) 次世代スーパーコンピュータ「CCC-5」の開発を(φが)進め、>

upper_of

<(4) (φが)来年末から64年初めにかけて(φを)出荷する>

方針を明らかにした。>

equal

<(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)。>

注) ・< , >は、単文の範囲を表わす。

・下線部は、フレームを表わす。

・φは、省略を表わす。

図-12 例文の照応処理結果の一部

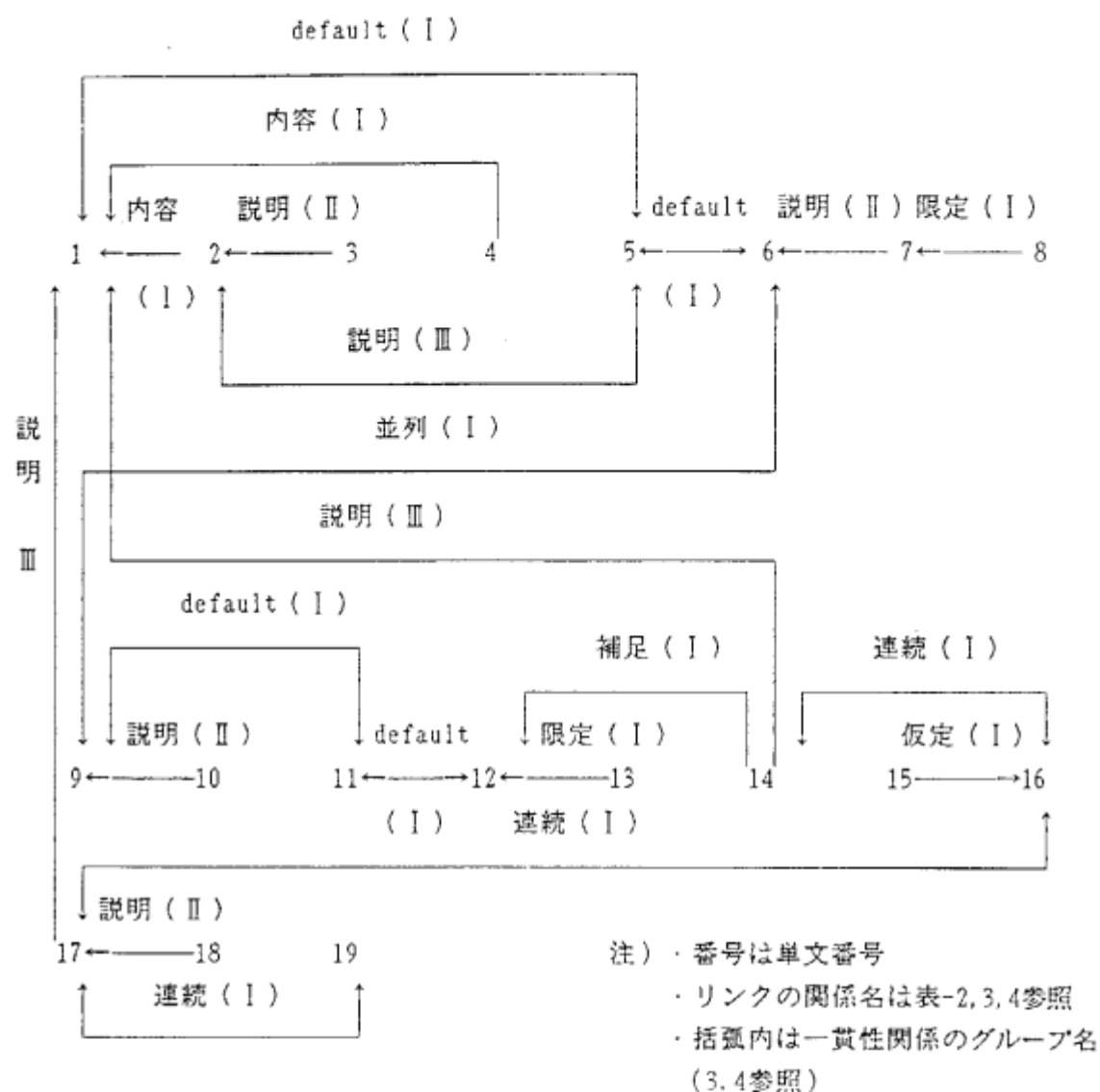


図-13 例文の一貫性処理結果

図米国のコンピューターメーカー、AAAAAAAAA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コンピューター「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1！0億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ーコンピューターで最も処理スピー！ドの速い米国DDDDDDDD社のEEEE-10Eモデル（6.9ギガFLOPS）の2倍以上とな！る。また、AAAAAAAAA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ーコンピューターは現在、約1！00台といわれ、うちAAAAAAAAA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、大阪で商談セミナーを行っている米国スーパーコンピュ！ター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-14 抜粋の例（評価値10以上の単語）

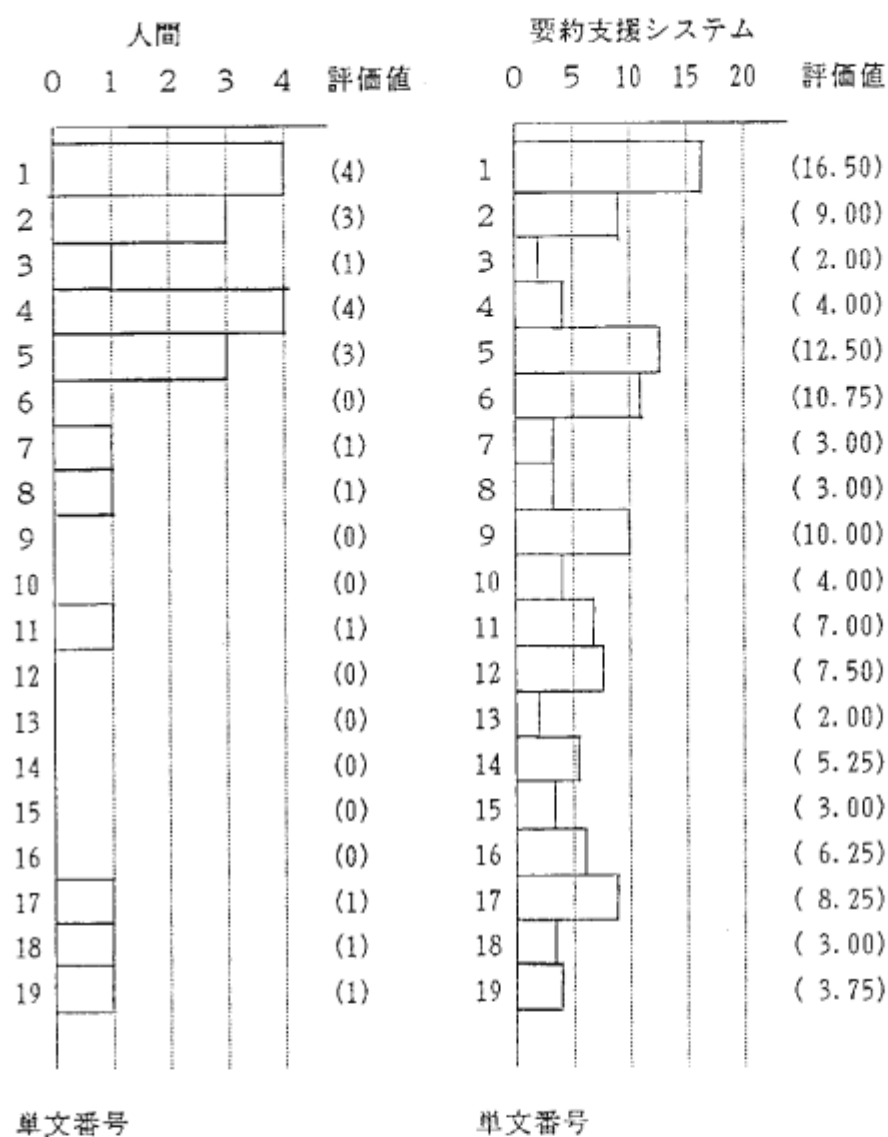


図-15 人間と要約支援システムによる重要性評価結果の比較

表-1 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）

表-2 1 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		<-->	2つの単文が連続した事象になっている
	並列		<-->	同題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		<-->	2つの単文間に他の接続関係がない

表-3 2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-4 3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の題目に照応関係がある
	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

