

文章のビジュアルな要約についての一考察

5F-9

田淵 篤、 原 良憲、 笠原 裕

日本電気(株) C&C情報研究所

1. はじめに

自然言語処理における文章要約¹⁾は、原文からより記述量の少ない文章への「フィルタリング操作」と見ることができる。しかしこの処理には、1)重要度の低い文を削除することで、文章としての連続性や機能的関連性などが崩れてしまう、2)文の重要度を判定する際にシステムが付加した文章の構造化情報(構文・意味)は、結局要約として表現されずに捨てられてしまう、などの問題点がある。これに対して、本稿で提案する文章の「ビジュアル要約」は、図表表現への「メディア変換」により、文章としての伝達内容をできるだけ減らさずに直感的な可読性を増すものである。すなわち、文章の論理的な構造や単語の指示対象を、概念図/表/アイコンなどで表現することで、文章を視覚的・2次元的に整理する。本稿では、文章をビジュアルに要約するための経験則を具体例を用いて論じる。さらに、この経験則を用いて計算機上で文章の解析結果から図表表現を再構築する手段について考察する。

2. 図表表現による文章要約

図表表現は、文字言語による記述と比べて、要素とその関係(因果・包含・対応など)を2次元に展開して包括的に表現するのに適している、という特徴がある²⁾。このような特徴を、文章を構成する各文の間に成り立つ論理的な関係付け(修飾的關係、参照關係、修飾關係など)の視覚化に応用することを考える。

これらの関係付けは、本来多次元のネットワーク構造を形成するものである。しかし、文章は文字の1次元的な並びであるため、そのような構造を表現しにくいことになる。関係を明示するためには主に代名詞や接続詞などの機能語が用いられるが、他の単語に比べて特に目立つものではないため、それらを拾い読みして論旨を把握するのは容易ではない。これに対して図表表現を用いれば、論理的な関係付けの構造をそのまま2次元的に提示できるので、論旨の概要をより直接的に表現することができる。このことは、KJ法におけるグラフ図やアウトラインプロセッサにおける章立ての階層構造図が、文章構造の整理の手法として用いられることからわかる。

3. 「ビジュアル要約」の例

「ビジュアル要約」において実際に用いられる表現技法を体系的に調査するために、工業デザイナー3人の協

力を得て、短い文章(15文程度)をビジュアルに要約する実験を行った。対象は比較的論理的構造のはっきりした説明文とし、「図表表現を用いて1枚のOHPにまとめる要領で」という示唆を与えた。図1a、bに原文とビジュアル要約の例を示す。

「××硝子」

製ビン業界第2位。主力のガラスビンが、ビールやジュースの用途を中心に缶・プラスチック容器に換わりつつあるため、PET樹脂、ニューガラスなど新規分野に注力中。

六四年三月期は、ドリンク剤向けガラスビン、清涼飲料向けPETボトルの二桁成長を背景に、売上高五二〇億円(四〇%増)、経常利益四四億円(五五%増)と大幅増収益を達成する見通し。PETは今期売上五〇億円規模に達し損益黒字化が定着しよう。

五九年よりニューガラスの研究に着手している。また、船底塗料の水溶性ガラスをサンプル出荷中で、高付加価値品展開の足がかりとなろう。

図1-a ビジュアル要約の対象とした文章 (文化創作出版・相場四季報'88年秋季号より)

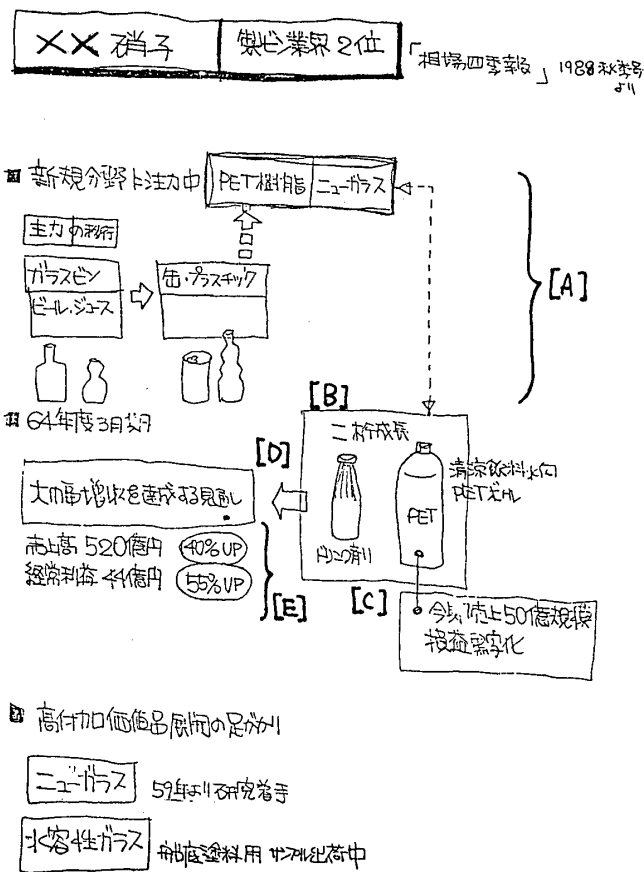


図1-b 作成されたビジュアル要約の例 ([A]~[E]は筆者による記入)

次に、図1-bを含めて今回の実験で得られたビジュアル要約の事例と照合しながら、文章の論理的構造がどのように図表表現の選択やレイアウトに反映されているかについて考察した。ここでは文章を構造化する要因として特に重要な次の3点に着目する。

- 文の中心的な記述対象（主題）の意味的な同一性によるまとめ（主題-題述の関係）
- ある文で言及された対象を別の文で再度言及することによる結束（参照関係）
- 各文の修辭的な機能を組み合わせることによる論旨の展開（修辭的關係）

1) 主題-題述の關係

文章は、主題についての記述（題述）によって、主題毎の小部分（文/パラグラフ）にまとめて構成される。図表表現においては、

- [A] 題述部分が比較的多い場合、主題を見出し（左上方）とし、題述の部分を見出しに近接する領域に配置する
- [B] 題述部分が1文以内で納まる場合、主題あるいは題述を見出しとする囲みの中に両者を併置するという手法が採られている（[A][B]は図1-bの参照記号に対応する。以下同様）。いずれにしても、構成要素を距離的に近づけて置くことで緩やかな関連付けを行っている。

2) 参照關係

参照關係は、文章中では代名詞などの参照のための機能語を用いて示される。しかしこのような間接的な表現では、参照されている対象を読者が文脈的な知識を用いて推論する必要がある。一方図表表現では、線分の持つ構成要素の連結作用（両端にある要素を強く関係付ける）を利用して、

- [C] 線分を引っ張って参照される要素と参照している記述とを接続する
- という、より直接的な表現が採られている。

3) 修辭的關係

修辭的關係には様々な種類のものがあるが、ここでは「因果關係」と「例証」の2つに絞る。因果關係は、事物の変遷や影響を表すものであるから、「流れ」のアナロジーで理解しやすい。文章中では「ので」などの機能語で示されることが多いが、図表表現では、

- [D] 因果關係にある事象を枠で囲み、枠同士を矢印で連結して並べる

という、いわゆる「流れ図」として視覚的に表現されている。例証は、具体例による例示が一般的な主張に比べて付随的な情報であることから、比較的弱く表現して關係自体が目立たない方がよい。そのため図表表現でも、

- [E] 一般的な主張の方を枠で囲んで強調し、具体例をその横に併置する

という方法が採られている。

3. ビジュアル要約の實現形態

本研究の最終的な目標は、2. で列挙したような文章の論理的構造を個別の図表表現に対応付ける規則を明確化し、計算機がその規則を適用することで、入力した文章に対するビジュアル要約を自動生成させることである。以下ではその實現形態について述べる。

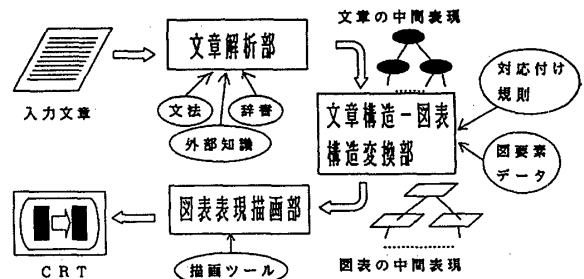


図2 文章のビジュアル要約システムの構成

システムの概念的な構成は、図2に示すような3つのモジュールから成る。

1) 文章解析部

入力文章を、形態素解析→構文・意味解析→文章構造解析という処理を経て、文章構造の中間表現に変換する。ここでは、単語の指示概念を最小単位とする概念の依存構造に変換することを想定する。

2) 文章構造-図表構造変換部

1) で得られた文章構造の中間表現を、図表表現の中間表現に変換する。ここでは、図表の構成要素をオブジェクトで、要素の包含関係をオブジェクトの階層関係で表現する。変換のための対応付け規則は、2. で列挙したような規則を

(IF 変換の対象となる文章構造が「因果關係」
THEN 01=前件の文、02=後件の文、03=矢印
01>03>02) ;;; 0i=図表の構成要素

のように形式化したものを用いる。この規則を適用することで、図表の構成要素の選択/配置に関する制約（規則のTHEN部）をオブジェクトの構造中に蓄積していく。規則適用を再帰的に繰り返すことによりオブジェクトは木構造に展開され、描画すべき図表表現の構成が図要素に至るまで漸進的に規定される。

3) 図表表現描画部

2) で得られたオブジェクトの構造を元に、実際に画面上に描画を行う。描画される内容、大きさ、局所座標はオブジェクトのなかで決定されているとし、階層構造の頂点から順に包含される画面領域にマッピングする。

4. おわりに

文章の論理的構造を図表表現を用いて2次元・視覚的に整理する「ビジュアル要約」の手法について提案した。現在筆者らは、解析済みの文章構造を対応付け規則に基づいて図表表現に変換するためのシステムを試作中である³⁾。今後は、• 対応付け規則の収集と形式化、• 図表表現の選択/配置に対する制約からの望ましい図表表現の描画、という点について実現を図っていく。

[参考文献]

- 1) 安原、小松、日比、加藤：要約支援システムCOGIT、情報処理Vol.30 No.10、p1258-1267、1989
- 2) 出原、吉田、渥美：図の体系、日科技連、1986
- 3) 田淵、原、笠原：メディア表現に対する制約に基づいた図表・文章の生成、第3回人工知能学会全国大会予稿集、p475-478、1989