自然言語処理 98－9
（1993．11．19）

語彙的結束性に着目した文章抄録法の提案

佐々木 一朗＊　　増山 繁＊　　内藤 昭三＊＊
{sasaki@smlab. , masuyama@}tutkie.tut.ac.jp*　naito@atom.ntt.jp**
＊豊橋技術科学大学知識情報工学系
＊＊ＮＴＴ基礎研究所
＊〒441, 愛知県豊橋市天伯町雲雀ヶ丘1-1
豊橋技術科学大学知識情報工学系 増山研究室
TEL. 0532-47-0111 (ext. 893)
＊＊〒180, 東京都武蔵野市緑町3-9-11

梗概

本稿では，日本語テキストの語彙的結束性表現法として提唱した，結束チャートを応用した文章抄録についての提案と，それによる結果，及び，被験者による抄録結果との比較について論ずる．結束チャートとは，テキストの各段落内，あるいは複数の段落に連続して出現する意味分類を，シソーラスに基づいて解析し，その結果をチャート形式で視覚的に示したものである．結束チャートの構成パターンから意味分類の重要度を算出し，その情報を手がかりとして抄録を行なう．実験テキストには，日経サイエンスの記事を使用した．

Lexical Cohesion based Extraction of Key Sentences

Ichiro SASAKI*　Shigeru MASUYAMA*　Shozo NAITO**
*Dept. of Knowledge-based Info. Eng., Toyohashi Univ. of Tech.
**NTT Basic Research Laboratories
*1-1, Hibarigaoka, Tenpaku-cho, Toyohashi-shi, Aichi, 441, Japan
*3-9-11, Midori-cho, Musashino-shi, Tokyo, 180, Japan

Abstract

We propose a lexical cohesion based extraction method of key sentences, using the cohesion chart proposed for analyzing lexical cohesion of Japanese texts, and discuss its effects and compare the results with those obtained by testees. The cohesion chart is constructed using the thesaurus of classifying words into large, middle and small categories which continue to appear among paragraphs in Japanese sentences. The proposed method selects key sentences in a text (an article in NIKKEI SCIENCE) based on calculating the importance level of sentences using the cohesion chart's structural pattern.

1 はじめに

大量の書籍，科学技術文献等の情報が溢れているが，その中から欲しい情報を的確に検索するのは困難な作業である．検索を補助するための情報として，キーワードや抄録の利用は有効な手段であり，そのための研究も精力的に行なわれている．本報告で提案する結束チャートを用いたテキスト抄録の目的は，一般の文章や技術的文献，論文等，幅広い分野の抄録の自動作成を実現することである．

一般に，「抄録」と「要約」とは，意味的にしばしば混用されている．辞書 [6] には，「抄録 = 抜き書きすること」，また要約についての定義は，「要約 = 大意を短くまとめること」とあり，いわゆる原テキストを短くまとめたものと考えられる．このように抄録と要約は，はっきり分けて考えられる．本稿で行なった抄録とは，そのテキストのあらましを述べている重要な文を幾つか拾い上げること，と考えることができる (本稿の抄録の詳しい定義は 3 章で行なう)．本稿では，抄録の自動生成に論点をおいている．

現在までに，自動抄録に関しては，水谷が [15] の中で，特に統計的な自動抄録法の定義について述べており，「何らかの統計的性質を利用して本文や主題との関係度合いの深い文を適当数だけ選びだし，そのトピックセンテンスを本文に現れた順に並べて抄録にする」という，[18] の定義を紹介している．また要約を含めた研究では，以下のように，2 つに大別できる．一つは，単語の頻度，要約的表現，キーワード，テキスト内の位置などを用いて要約を行なう，いわゆる文字処理などによる表層的な方法 [10, 14, 15, 18] である．特に，[18] により提唱された，キーワード密度方式による抄録作成の方法は，日本語への応用が試みられ，[15] によって基本的な問題点が指摘され，改良が行なわれている．その後も更に，[10, 14] によって改良が行なわれている．

もう一つの方法は，スクリプトに基づいた方法や物語文法を使用してテキストを解析し，重要箇所や事項を抽出するトップダウン的方法や，ボトムアップ的にテキストを解析し談話構造を意味ネットワーク等を用いて意味表現を行ない，重要箇所を決定する方法など，いわゆる構文，意味処理による方法 [5, 7, 9, 11, 12, 13, 17] である．しかし，これらの研究では，ある特定の分野や限られた内容についてのみ書かれたテキストを対象としている場合が多く，汎用性に欠ける面もある．

本稿では，すでにテキストの語彙的結束性の表現法として提唱した [1, 2] 結束チャートと呼ぶデータ構造を用いた抄録法を提案し，実験を行なう．本抄録法は，上記の分類に対応させると，表層，統計的な方法と意味処理による方法を併用したものとみなすことができる．結束チャートを用いることで，段落内，又は，段落間に跨る意味分類の出現パターンを視覚的に容易に把握することができるので，テキスト全体の語彙的結束性に関する大域的構造を反映でき，抄録作成の際の意味的，また文章の全体的な構造の情報源として有効である [1, 2]．この特徴を活用し，計算機上でこの特徴を捕らえる計算を行なうための定式化を行ない，意味，分類的な抄録情報として使用するための一方法を紹介する．今回の抄録実験では，科学技術文を使用したが，本手法は汎用的であり，対象テキストの分野を限定する必要はない．更に，実際に数人の被験者に文と段落の抄録を行なってもらい，結束チャートを用いた自動抄録実験との比較を行なうとともに，人間の抄録における傾向を分析した．以下，2 章では結束チャートの概要を紹介し，3 章では抄録法を提案し，抄録実験システムについて説明する．4 章では，抄録アンケート調査の方法とその結果を述べ，5 章では結束チャートを使った抄録実験結果を提示し考察を行なう．むすびでは，今後の課題などを述べる．

2 結束チャート

結束チャートとは，テキストの各段落 (又は意味上のひとまとまりや章など) 内，あるいは複数段落に連続して出現する意味分類を，シソーラスに基づいて解析し，チャート形式で表示したものである．

この結束チャートを作成する場合に，その構築法としていくつかの考えられる方法のうち，テキスト頭から現在処理を行なっている位置までの中分類の中で，累積頻度 (総合頻度) が最大の中分類を選択する方法を採用した．ただし，初めて出現する分類などの理由により，タイが起こる場合は，若い番号の分類を選ぶという方法で行なった．この方法を選択した理由については，詳しくは [1, 2] の中で，結束チャートの作成方法について示しているので，そちらを参照されたい．

2.1 シソーラス

シソーラスには，「角川類語新辞典」[6] (収録語彙数 57145 語) を用いた．この辞典では各語に 10 進 3 桁の意味分類番号が付されていて，それぞれ，大，中，小の各分類に対応している．本研究では，結束チャートの生成には中分類を使用し，分類の頻度統計の対象には，小，中分類を使用している．またシソーラスは，形態素辞書としても併用している．

2.2 使用テキスト

抄録実験テキストには，日経サイエンスの記事 3 編を使用した．タイトルは，それぞれ，「もう 1 つの太陽系を探す (TEXT 1) 文数 203, 小段落数 52, 大段落数 6」，「エネルギー貯蔵システム (TEXT 2) 文数 255, 小段落数 55, 大段落数 11」，「これからのクルマはどうなるか (TEXT 3) 文数 264, 小段落数 81, 大段落数 10」であり，全く異なった主題に関する内容のものである．日経サイエンスを選んだ理由は，充分練られたテキストであることが

挙げられる．日経サイエンスの結束チャートは，段落を単位として作成したが，この場合の段落とは，幾つかの小段落 (一字下げが行なわれた形式段落) からなる，小題の付されている大段落を意味している．日経サイエンスにおいて結束チャートの作成単位として大段落を採った理由は，次の通りである．

(1) 小段落では一テキスト中の段落数が多くなり過ぎるため，チャートの持つ一覧性などの性質が損なわれてしまう，

(2) 小主題によってまとまりをもった単位に対して作成することにより，結束チャートの談話構造表現力を有効に利用できる．

3 自動抄録法

本稿の抄録実験のねらいは，文章の大域的構造を表す結束チャートが抄録生成に有効であることを検証することである．ここで，抄録とは，1 章で述べた定義を拡張して新たに，

「テキストからその文章全体の内容，又はタイトルなどから判断して，そのテキストの内容全体の意味を捕らえていると思われる文，または小段落を，テキスト中から幾つか抜き書きしたもの」

と定義する．

タイトルを抄録作成の判断基準のひとつとした理由は以下の通りである．

- 木下は，レゲット [3, 4] を引用して日本語ではいくつかのことを書き並べる時，その内容や相互の連関がバラグラフ全体を読んだあとで始めてわかるというような書き方がなされていると述べている [8]．このような談話の展開は，中央に一本の幹の通った逆茂木型もしくは樹木型の樹 (レゲットの樹とよぶ) の構造となり，幹が首尾一貫性を構成すると論じている．このような構造は，結束チャートにも表現される．タイトルは，首尾一貫性を構成する幹の情報を表すものと考えることができる．

定義にも述べたように，今回は，文と小段落を単位として抄録を行ない，抄録における文と小段落の関係を明らかにすることを試みた．自動抄録結果を評価する比較データを得るために，筆者の研究室内の学生 16 名にアンケート調査を行なった．アンケート調査では，自動抄録を行なったテキストと同じテキストについて，文と小段落を単位とする抄録を行なってもらい，両者の抄録結果を比較考察した．アンケート調査の結果は 4 節で述べる．

抄録処理の概要は以下の通りである:

1) 形態素解析
2) 単語の意味分類の抽出
3) 結束チャートの作成
4) 結束チャートに基づく意味分類の重み指数付け
5) 各文，段落単位の重み付け
6) 抄録となる文，段落の抽出

3.1 システム構成

抄録システムの概要を以下に述べる．システムは形態素解析部，結束チャート作成部，結束チャート情報作成部，抄録情報計算部の 4 つのモジュールからなる．各モジュールの処理概要は以下の通りである．なお，結束チャート作成部の詳細は，文献 [1, 2] を参照して欲しい．

形態素解析部 ([1, 2] のものを一部変更) テキストからの辞書中の語，意味分類情報，及び，未登録語の中での特徴語 (仮名文字や，カッコなどの記号で囲まれた語) を抽出する．形態素解析用の辞書には，シソーラス (角川類語新辞典) と ICOT で開発されたもの (フリーソフトウエアとして提供されている TRIE 辞書ユーティリティ中に含まれているもの) を組み合わせて使用している．形態素解析プログラムは，C 言語で作成した最長一致法による形態素解析プログラムと，ICOT 作成のフリーソフトウエア中のプログラムとを組み合わせて使用している．形態素解析の段階では，複数の意味分類を持っている語 (以下，多義語と呼ぶ) に対しては，意味分類の選択を行なわず，すべての意味分類を与えておく．シソーラスに掲載されていない語 (以下，未登録語と呼ぶ) の意味分類の処理は行なわなかった．未登録語の処理を行なわない理由は，シソーラス中にない情報を加えることは行なわず，あくまでもシソーラスに対して客観的な結束チャートを作成するためである．

結束チャート作成部 [1, 2] 形態素解析部からの情報から，多義語の意味分類の決定を行ない，結束チャートデータを作成する．

結束チャート情報作成部 結束チャートの形状から各中分類毎に分類の重み付けを行なう．詳しくは 3.2 節で説明する．

抄録情報計算部 形態素解析結果から計算される語の頻度情報と各語の意味分類に関する結束チャート情報に基づき，一文の重みを計算し，重みの大きいものから順に抄録文あるいは小段落を選択する．

抄録情報計算部での語の頻度情報は，未登録語は全て名詞とみなしたうえで，未登録語も含めた語の頻度情報を計算している．

3.2 結束チャートの使用方法

本節では，結束チャート情報作成部の詳細を説明する．この部分での処理の目的は，結束チャートの形状から，各中分類に対し重み指数 (重要度を示す指数) を計算し，後続の文の重み計算に必要な情報を作成することである．重み計算は以下の方法で行なった．まず，以下の式により，各意味分類に

対する段落ごとの重みのプラス要因，マイナス要因を計算する．

P_k = (その段落kからの出現段落長/全段落数) $*$ g +(その中分類における全出現分類数)/120 .(1)

M_k = (その段落kからの出現段落長/全段落数) $*$ m .(2)

P_k：各段落中でのプラスの重みの和

M_k：各段落中でのマイナスの重みの和

ここで，パラメータ，g, m, は各意味分類の出現する段落数に依存して決まり，今回の実験では，表1の値を使用した．ただし，表中のNとは，全段落数を示す．

出現段落長	g	m
N	2.0	1.5
N−1	1.8	1.4
N−2	1.6	1.3
4以上N−2未満	1.5	1.3
3	-	1.2
2	1.2	1.1
1	1.0	1.0
0	0	0

表1. プラスの重み g, マイナスの重み m の値

つぎに，全段落に対する各意味分類の重みは以下の式で計算される．

$$G = \sum_{k=1}^{\text{全段落数}} P_k - \sum_{k=1}^{\text{全段落数}} M_k \quad .(3)$$

G：その中分類の重み指数

最後に各文に対する抄録候補としてのもっともらしさを示す指数Wの計算は次式による．

$$W = \left(\sum_{\text{文頭}}^{\text{文の終り}} (n * G) * (1 + log(1 + \text{中分類頻度}) + (1 + log(1 + \text{小分類頻度})) + (1 + log(1 + \text{単語頻度}))) / \text{文の長さ} \right. \quad .(4)$$

n：形態素解析で区切られた各語

式の意味を簡単に説明する．(1), (2) 式は各分類の連続する出現，不出現長のチャート出現パターンにより，重みを計算する．これらの式を使うことにより，余分な分類は重みが小さくなり，必要な分類が特徴的に大きな重みをとる．(1) 式において，出現中分類数を120で割った理由は，一文章中での中分類の最大出現数を120とみなしたためである．パラメータ m,g の値は，各意味分類の段落に対する出現，不出現長さに対する重み付けである．(3) 式では，(1), (2) 式の全段落での和をとり，(4) 式では，各分類，単語の頻度を加えることにより，各文の重要度を計算する．最後に，各頻度に対し log をとった理由は，各頻度の影響を小さくして結束チャートの構造的特徴をできるだけ大きく反映させたいと考えたためである．

4 アンケートによる抄録調査

結束チャートを用いる抄録結果を評価するための比較対象となるサンプルを収集するため，並びに，今回用いたテキストに対する人間の抄録パターンを解明するために，研究室内の学生を被験者として抄録アンケートを行ない，16人から回答を得た．調査は以下の2つの質問項目について，3テキストを対象に (TEXT 1, 5人,TEXT 2, 5人,TEXT 3, 6人) 行なった．

調査1.の質問 テキスト中から，タイトルの意味に沿っていると考えられる文を，最も的確に表現していると考えられるものから順番に20文以上30文以内の範囲で抜き出して下さい．

調査2.の質問 タイトルの意味に沿った小段落を10個抜きだし，最も意味に沿った小段落を上から順番にソートして並べて下さい．

質問1, 質問2のどちらを最初に行なうかの指定はしなかった．そのため，文と小段落の選び方に一部相関が見られた．ほとんどの人は，文と小段落を独立に選んでいたが，一部の被験者は，小段落を選び，その中から文を選んでいくという方法を採用した．アンケート調査の結果に見られた特徴を以下の図，表に示す．なお，大段落番号は，文章の始めから1, 2, . と順に付している．

4.1 文抽出位置調査

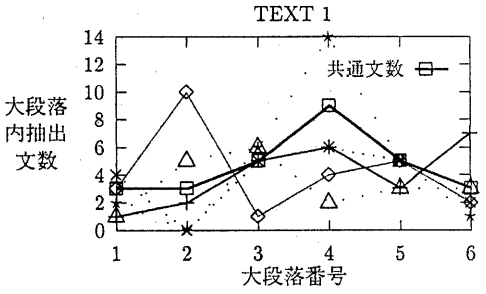

図1. 各大段落内抽出文数

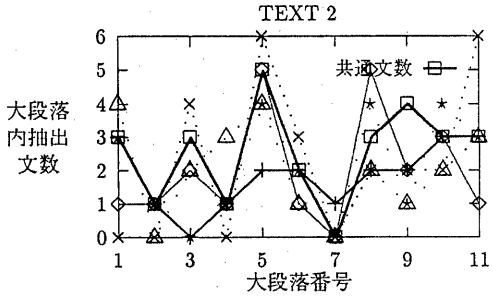

図2. 各大段落内抽出文数

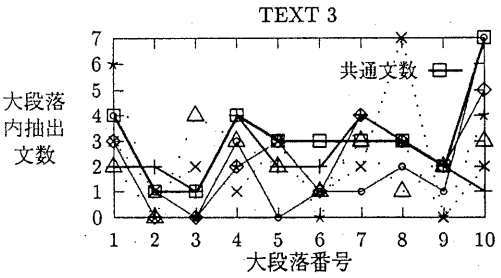

図 3. 各大段落内抽出文数

注) 各図中の線は，それぞれ被験者が採った抄録文の数を各大段落毎に示しており，共通文数とは，各被験者が抽出した各大段落内の文のうち，同じ文を複数人が選んだ文数を大段落毎に示した数である．

テキスト	選択割合平均 (%)
TEXT 1	50.1
TEXT 2	47.7
TEXT 3	52.2

表 2. 抽出した小段落内に抽出文が含まれていた割合

4.2 小段落抽出位置調査

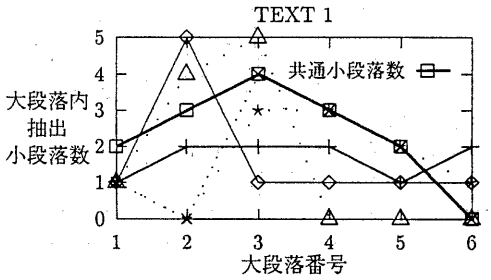

図 4. 各大段落内抽出小段落数

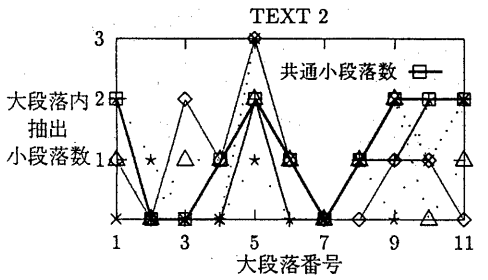

図 5. 各大段落内抽出小段落数

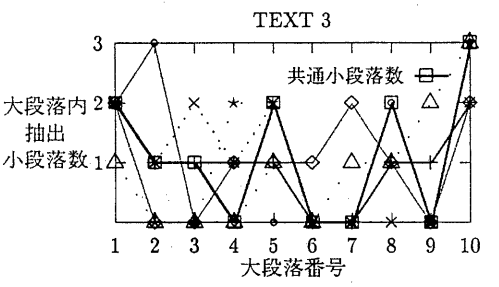

図 6. 各大段落内抽出小段落数

注) 各図中の線は，それぞれ被験者が採った抄録小段落の数を各大段落毎に示しており，共通小段落数とは，各被験者が抽出した各大段落内の小段落のうち，同じ小段落を複数人が選んだ小段落数を大段落毎に示した数である．

4.3 抽出文及び小段落調査

それぞれ，被験者が選んだ文，小段落のうち，同じ文，小段落が選ばれた場合の選んだ人数分布を以下の表に示す．但し，正解抄録数とは，2 人以上に選ばれた各文，小段落の事であり，これら文，小段落を結束チャートによる結果と比較する．

共通文選択を人数	TEXT 1 () 内は %	TEXT 2 (%)	TEXT 3 (%)
全員一致	0 (0)	2 (3.1)	2 (3.0)
5 人	-	-	2
4 人	4 (5.1)	4 (6.2)	7 (10.6)
3 人	9 (11.5)	7 (10.8)	5 (7.6)
2 人	14 (17.9)	16 (24.6)	15 (22.7)
1 人	51 (65.4)	36 (55.4)	35 (53.0)
全抽出文数	78	65	66
正解抄録数	27	29	31

表 3. 共通文選択人数とその文数

共通小段落を選んだ人数	TEXT 1 () 内は %	TEXT 2 (%)	TEXT 3 (%)
全員一致	1 (4.7)	0 (0)	1 (4.1)
5 人	-	-	4 (16.7)
4 人	3 (14.3)	3 (11.1)	0 (0)
3 人	0 (0)	4 (14.8)	0 (0)
2 人	9 (42.9)	6 (22.2)	4 (16.7)
1 人	8 (38.1)	14 (51.9)	15 (62.5)
全抽出小段落数	21	27	24
正解抄録数	15	13	10

表 4. 共通小段落を抄録として選んだ人数とその小段落数

4.4 アンケート調査による抄録の考察

アンケート調査結果から人間による抄録パターンの傾向について考察する.

- 実験対象テキストがかなり長いので, タイトルの意味に沿った文, または小段落を抜きだしてから更に, 重要と思われる順にソートしてもらうというアンケート調査の結果, ソートの順番は, 下位のものはバラツキが大きく, 逆に, 上位に選択されていた文, 小段落については, 被験者によらず大体の一致が見られた.

- テキストに書かれている内容が主観的か客観的か, また, タイトルの付け方の適切度に応じて, 抽出文の種類 (ex. 説明, 意見, 推測文, 他等) の選ばれ方に偏りが見られた.

- 特に TEXT 2,TEXT 3 に対しては, 各 2 ～ 3 人ほどの被験者から抄録が難しかったという意見が出された. その理由として, 原文に付されているタイトルがそのテキストにふさわしくないという意見が 2 名ほどの被験者からあった.

- 各大段落中から選ばれた共通の文, 及び小段落の抽出傾向 (各図において太い実線で示されている線) は, ほぼ同じような形となっている. これは, 各テキスト中の大段落で抄録として興味をひく部分が一致していることを表していると見られる.

- 抽出文, 小段落とが一致している割合は表 3 に示している通り 50% 前後であり必ずしも文と小段落とが一致するとは限らないことが明らかとなった.

- 文章構成上の起承転結に従い, 段落の最初と最後に重要な文が比較的多く現れるということを示す, 抄録抽出位置分布における傾向が現れていることが観察される. この傾向は, 特に TEXT 2,TEXT 3 について顕著である.

- 文章全体から平均して抄録文が得られるというよりも, むしろ, ある特定のまとまりを持った部分 (今回の調査での大段落に当たる) から偏って抄録されるということが観察される. このことは, 各被験者にとって印象的な部分 (今回はタイトルの意味に沿っているということ) が共通していたこと, あるいは, 注視点[1]が共通していたことを示していると考えられる.

[1] [16] 参照.

5 結束チャートによる抄録結果

今回の結束チャートによる抄録の結果を以下の表に示す. 前章の調査より得られた結果で, 1 テキスト中, 文, 段落ともに 2 人以上の一致があった文, 小段落を, 正しい抄録 (以下, 正解文, 正解小段落, 両方を示したものを正解抄録と呼ぶ) とみなし, 結束チャートによる抄録と比較した結果を再現率と適合率により表している. ここで, 再現率とは, 全正解数に対する抽出正解数の割合, 適合率とは, 全抽出数に対する抽出正解数の割合であり, 以下の式で定義される.

$$\text{文の再現率} = \frac{\text{自動抽出した抄録文に含まれる正解文数}}{\text{正解文数}}$$

$$\text{小段落の再現率} = \frac{\text{自動抽出した抄録小段落に含まれる正解小段落数}}{\text{正解小段落数}}$$

$$\text{文の適合率} = \frac{\text{自動抽出した抄録文に含まれる正解文数}}{\text{抽出文数}}$$

$$\text{小段落の適合率} = \frac{\text{自動抽出した抄録小段落に含まれる正解小段落数}}{\text{抽出小段落数}}$$

以下の表の適合率のデータ中に付された数字は, 自動抄録において, 上位いくつまでを候補として抽出したかを示したものである. 再現率については, 正解抄録数として得られた数までを抄録候補の範囲としたものを掲載している.

	TEXT 1	TEXT 2	TEXT 3
再現率	29.6	13.8	25.8
適合率 30	26.7	13.3	26.7
適合率 20	35.0	15.0	20.0
正解文数	27	29	31
総文数	203	255	264

表 5. 文の再現率と適合率

	TEXT 1	TEXT 2	TEXT 3
再現率	53.3	38.5	30.0
適合率 15	53.3	33.3	26.7
適合率 10	60.0	40	30
正解小段落数	15	13	10
総小段落数	52	55	81

表 6. 小段落の再現率と適合率

5.1 結果の考察

まず, 結束チャートを用いた抄録結果について考察する.

A1. 自動抄録の結果において, 正解として採った範囲 (適合率で, 自動抄録で上位幾つまでをその抄録の正解としたかという範囲) から, 離れていても, その部分でいくつかまとまって抄録文として採用されていることが見られた. 特にこれは, 文を単位とした抄録結果に多く見られた.

A2. 重要度計算に関するパラメータ (重みの m,g) を変更することで, 大きく結果が変わってくることが実験の結果明らかとなったので (筆者がかんじた, ある意味的分類 { 今回用いたシソーラスの分類に特に限った意味ではなく } について, その分類毎にまとまった形で出現する), これを利用してある特定の意味的な情報抽出の可能性を検討する必要がある.

A3. 自動抄録候補は, ほぼタイトル通りのものが選択されており, あまり関係の無いと思われるものは順位が下位の方にあった.

次に, 結束チャートを用いた抄録と抄録アンケート結果との比較について考察する.

B1. A3 とも関連があるが, 適合率でとった正解範囲の比較による部分の結果から, 全体的に上位に正解抄録が得られていることがわかる. これは, 今回の正解抄録が, そのテキスト中で何らかの意味的, 分類的な特徴があったと思われる.

B2. 全体的に再現率, 適合率ともに低いが, 小段落単位の抄録の再現率, 適合率が, 文単位の抄録のそれらに比べて若干勝っている. この理由としては, 文単位の抄録を行なうためには, 結束チャートのようなテキストの大域的意味情報や, 丸め込んだ語や分類の頻度情報のみでは文単位の情報量が少なくなることが考えられる. 文単位の抄録の精度を上げるためには, 格情報などの詳細な情報も併用する必要があると考えられる. また, 頻度情報 (語や分類の) を結束チャートにとり込むことにより精度向上が可能になると考えられる.

B3. 小段落単位の抄録において, TEXT 1 の結果は, TEXT 2, TEXT 3 に比べて明らかに優れている (表 6 参照). この結果は, 中にあった「TEXT 1 の抄録はあまり難しくはなかったが, TEXT 2, TEXT 3 については, 抄録を採るのが非常に難しかった」というアンケートに関する何人かの被験者の意見を裏付けるものである.

B4. 結束チャート自体が, 段落毎の大域的構造を表示するという性質を持っているので, 今回の抜き書きをするという, 抄録に関しては, 文単位よりも小段落単位の方が精度良く行なわれたものと考えられる.

6 おわりに

本稿の結束チャートを応用した抄録は, 抄録情報抽出の手段として有効なものであるということが, 今回の実験から明らかになった. 抄録が, かなり各個人レベルで受ける印象 (今回はタイトルの意味に沿った抄録とした) により結果が異なって出てくるということが改めて明らかとなった.

今後の課題としては, 以下のようなものがある.

1. テキストの数を増やし, 自動抄録で抽出した結果について, アンケート等でそのふさわしさを点数付けしてもらい, この自動抄録結果を確かめるという方法により評価を行なう.

2. 語の頻度や分類に対する頻度を少し上げたりするなどして, 実験を重ね, よりよい情報がさらに得られるよう試みる.

3. 人間による抄録の傾向調査と, 結束チャートを用いた抄録についての実験データを更に検討考察し, 今回用いた方法での問題点を洗い直し, 更に現実的な抄録が得られるよう試みる.

4. 人間が受けるテキストの印象をうまくチャートから再現出来るような定式化について, さらに再考する.

5. この結束チャートを用いた抄録を要約等を行なう際の前処理的な段階での, 重要な段落, 部分等の絞り込みを行なうという部分で使用することを検討する.

6. 今回は抄録に重点を置いたが, 実験を行なっている中で同時に, ふさわしいタイトル付けに関しても, 結束チャートの利用可能性への糸口が発見できたので, 今後検討を行なう.

謝辞

「角川類語新辞典」を計算機可読辞書の形で提供していただき, その使用許可を頂いた (株) 角川書店に深謝する. またアンケートによる抄録に協力頂いた阪田, 増山研究室の学生の皆様に感謝する.

参考文献

[1] 佐々木一朗, 増山繁, 内藤昭三 ： 結束チャートを用いた日本語文章の語彙的結束構造の解析, 情報処理学会第 46 回全国大会講演論文集 (3) ,pp.3-177-178 (1993).

[2] 佐々木一朗, 増山繁, 内藤昭三 ： 結束チャートの自動生成と日本語文章の語彙的結束構造解析への応用, 情処研究報告 NL-95-8, pp.57-64 (1993).

[3] A. J. Leggett ： Notes on the Writing of Scientific English for Japanese Physicists, 日本物理学会会誌, 21, (1966) 790.

[4] A. J. Leggett ： Notes on the Writing of Scientific English for Japanese Physicists, 「Journal の論文をよくするために」, 増訂版, 日本物理学会, pp.96 (1975).

[5] 稲垣博人 ： 事象解析による要約情報の抽出, NLC91-9, pp.17 ～ 24, (1991).

[6] 大野晋, 浜西正人 ： 角川類語新辞典　角川書店 (1981).

[7] 小川泰嗣, 望主雅子, 別所礼子 ： 複合語キーワードの自動抽出法, 情処研究報告 NL-97-15, pp.103-110 (1993).

[8] 木下是雄 ： 理科系の作文技術, 中央公論社 (1981).

[9] 木本晴夫 ： 日本語新聞記事からのキーワード自動抽出と重用度評価, 電子情報通信学会論文誌, Vol.J74-D-I,No.8, pp.556-566, Oct, (1991).

[10] 鈴木康広, 栃内香次 ： キーワード密度方式自動抄録法の改良, 情報処理学会論文誌 Vol.29,No.3, pp.325 ～ 328, (1988).

[11] 住田一男, 小野顕司, 三池誠司 ： 対話的文書検索のための文書構造解析, 情処研究報告 NL-97-11, pp.71-78 (1993).

[12] 田村直良 ： 要約過程の形式化と現実について, 人工知能学会誌 Vol.4, No.2, pp.196 ～ 206, 1989.

[13] Hashida, K. Ishizaki, S., Isahara, H. ： An Approach to Abstract Generation, Bul. Electrotech. Lab., Vol.52,No.4, pp. 551 ～ 564, (1988).

[14] 間瀬久雄, 大西昇, 杉江昇 ： 説明文の抄録作成について, NLC89-4, pp.5 ～ 12, (1989).

[15] 水谷静夫 ： 統計的自動抄録法の問題点, 計量国語学,27, (1963).

[16] 宮崎清孝, 上野直樹 ： 認知科学選書 1　視点, 東京大学出版会 (1985).

[17] 安原宏, 小松英二, 日比孝, 加藤安彦 ： 要約支援システム COGITO, 情報処理 Vol.30,No.10, pp.1258 ～ 1267, (1989).

[18] Luhn, H.P ： The Automatic Creation of Literature Abstracts, IBM JOURNAL159 ～ 165, April (1968).