

光と影を用いたテキストのテーマ関連度の可視化

佐藤 圭太^{†1} 西原 陽子^{†2} 砂山 渡^{†1}

近年、Web上で情報発信を行う機会が、飛躍的に増大している。しかし、Web上に存在する文章の多くは、文章を書くプロによるものではないため、必ずしも文章の主題と内容の一貫性が保たれているわけではない。そのため、主題と内容の関連の把握に、多くの時間と労力を要することがある。

そこで本研究では、文章の主題(テーマ)と各文の関連度を評価し、文章中の各文を光と影により可視化する、ひなたシステムを提案する。すなわち、ユーザが文章のテーマを文章のタイトル中の名詞の集合などによって与え、システムは、文章中のテーマに関連する文に光を、関連しない文に影を与えて出力する。これにより、文章においてテーマと関連する部分の効率的な把握や、文章の全体的なテーマとの関連を知ることが可能になる。実験により、本システムが、文章のテーマとの関連把握に有効であることを検証した。

Text Visualization using Light and Shadow based on Topic Relevance

KEITA SATO,^{†1} YOKO NISHIHARA^{†2} and WATARU SUNAYAMA^{†1}

There are so many opportunities to transmit text information using the World Wide Web. However, texts on the Web are not always written by a professional writer, therefore, those may not be coherent and may be hard to be comprehended. As a result, we should take too much time and energy to grasp topic relevance.

This paper describes HINATA System that visualizes texts using light and shadow based on topic relevance. Topic is defined as a set of words such as nouns contained in a text title. The light expresses sentences related to the topic, and the shadow expresses sentences unrelated to the topic. By using this visualization method, users can find the parts related to a topic, and know the relationship between a text and a topic efficiently.

Experimental results showed that the proposed method could support users to understand how a text was related to a topic.

1. はじめに

近年、World Wide Web(WWW)の発達により、個人が簡単に情報を発信、取得できるようになってきた。特に、メール、Webサイト、Weblog、電子掲示板を通じて、文章のやりとりをすることが、ごく当たり前になって来ている。最近では、個人の情報発信の手段として、YouTube¹⁶⁾ やニコニコ動画¹⁷⁾ など、動画をメインのコンテンツとしたWebサイトも注目されている。しかし、これらのWebサイトにおいても、動画へのコメント機能や掲示板機能を設けるなど、コミュニケーションの手段として文章を扱う点は外すことができない。

Web上に存在する文章の多くは、文章を書くプロではないごく普通の個人が書いたものであるため、必

ずしも、文章の主題と内容の一貫性が保たれているわけではない。そのため、理解が難しい文章や、意図がうまくつかめない文章、無駄に長い文章等が存在している。またプロが書いた文章であっても、長い文章を読むには労力がかかったり、読み手によって、読みたい部分が変わる可能性がある。

そこで本研究では、文章の主題(テーマ)と各文の関連度を評価し、文章中の各文を光と影により可視化する、ひなたシステムを提案する。文章のテーマと関係のある部分、および関係のない部分を可視化することで、テーマと関係のある箇所と、その全体に対する位置や割合を確認でき、テーマとの関係に基づいて文章の理解を支援できると考えられる。

以下、本論文では、2で関連研究、3でテキストに光と影を与えるひなたシステムの構成と詳細について述べる。4で本システムの与える光の妥当性を確認する予備実験について述べる。5で本システムの光と影が、テキストの内容把握に対して有効に働くことを確認した評価実験について述べ、6で本論文を締めくくる。

^{†1} 広島市立大学大学院情報科学研究科
Graduate School of Information Sciences, Hiroshima City University

^{†2} 東京大学大学院工学系研究科
School of Engineering, The University of Tokyo

2. 文章理解支援の関連研究

本章では文章理解支援の関連研究について述べる。

2.1 文章の表示方法と理解力の関連

文章の表示方法を工夫すると、その内容を理解しやすくなる。従来研究において明らかにされてきた。文章を読む際には、自らが重要と思う箇所に線を引くと文章の理解が進むことに加え、他人によって下線が引かれた箇所に注意して読むことでも文章の理解につながる。また、重要なキーワードを強調して表示することや箇条書きを用いること⁹⁾、文章を適切な段落で切り分けて表示すること⁸⁾、重要ではない文を削除して表示すること¹⁴⁾なども効果があると知られている。これらの研究に対し、本論文では文章に背景色をつけることによる、文章理解への効果を明らかにする。

これまでの研究でも文章に背景色をつけることにより、文脈変化を理解させる効果があることは知られている¹⁰⁾。これに対し本研究では、文章のテーマとの関連を背景色により表現することで、文章のテーマとの関連、内容把握を支援する。

2.2 文章構成の評価と可視化

文章の構成を評価して可視化するシステムが多数提案されてきており、その一つとして議論の議事録の構成を可視化し、議論支援を行うシステムがある^{2),5)}。例えば松村らのシステムでは、議事録を話題ごとに分割し、分割された議事録間の関連を評価して、議事録の構造を可視化することができる⁵⁾。また、小論文のように一人の人間によって書かれた比較的長い文章に対しても、複数のシステムが提案されてきている^{3),4),6)}。例えば石岡の手法では、小論文の接続詞と文末表現を用いて小論文の構成を評価し、評価結果を星座盤により可視化する³⁾。これらの研究と本研究は、文章構成の可視化という点で共通しているが、本研究では文章と文章の関連ではなく、テーマと文章の関連を評価する点で異なる。すなわち、文章のテーマとの関連を評価することにより、テーマに関しての文章理解を支援する。

文章の構成ごとに異なる色を用いて文章の背景に彩色し、文章の理解を支援するシステムが提案されている¹²⁾。このシステムで彩色を行うのはユーザであり、ユーザの主観に基づき彩色を行うことが可能になっている。この研究と本研究とは、文章の背景に彩色して文章理解を支援する点で共通しているが、本研究ではシステムが自動的に彩色を行い、またユーザは彩色パターンを決める単語を変更できる点でこの研究とは異なっている。すなわち、システムが彩色を行うことにより、半自動的に配色を行なうことで、配色のための労力を軽減できると考えられる。

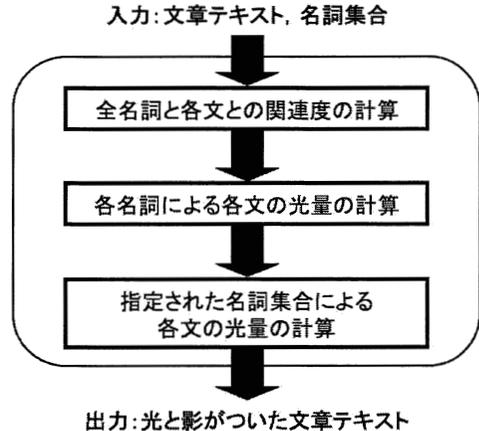


図1 ひなたシステムの構成

2.3 文章の要約

従来からの文章要約研究も、文章のテーマに関する理解を支援する。文章要約手法の多くは文章中の各文に対して重要度を与え、重要な順に文を抽出して出力する。重要度の評価方法としては、高頻度の単語が多数含まれるほど重要とするものや、テーマに関連する単語を多数含むほど重要とするものなどがある^{7),7),11),15)}。しかし要約システムの場合、要約として選ばれなかった文は出力に含まれることはない。提案するシステムでは、テーマとの関連が強い文ほど重要な文として光を当てて出力するが、重要でない文にも暗い光や、影を当てて出力を行なう。そのため、要約システムでは見落とされがちな文を確認できたり、重要な文のテキスト全体の中での位置を直感的に把握することが可能になると考えられる。

3. ひなたシステム

本章では、文章テキストと、名詞の集合（ユーザが考える文章のテーマ）を入力として、テーマに関係する部分を光、関係しない部分を影で表現する、ひなたシステム（図1）について述べる。以下で、本システムの詳細について述べる。

3.1 入力：文章テキストと名詞集合

本システムは、視覚化を行いたいテキストと、そのテキストに光を与えるもととなる名詞の集合を入力とする。

対象となる入力テキストは、句点等によって一文ごとに切り分けが可能なテキストで、かつ形態素解析によって名詞の抽出が可能なテキストとする。なお形態素解析には、「茶筌」²⁰⁾を用いた。

加えて、ユーザが考える文章のテーマを名詞の集合として与える。ユーザがテーマを指定せず、かつテキ

ストのタイトルが存在するときには、そのタイトルに含まれる名詞集合を入力とすることもできる。また、一度視覚化の結果を表示させた後に、再度、初めに与えた名詞集合とは異なる集合を与え、インタラクティブな視覚化を行うことも可能である。

3.2 全名詞と各文との関連度の計算

本節では、テキスト中の各文に、各名詞との関連度を与える方法について述べる。本モジュールでは、以下で述べる方法を、テキスト中の全名詞^{*1}に対して適用する。

ある名詞 *noun* に対する、*i* 番目の文の関連度 $Relate_i(noun)$ は、その文に含まれる、名詞 *noun*、およびその名詞の関連語の数として与える。

具体的には、名詞 *noun* の関連語集合 $Rset(noun)$ を元に、以下のアルゴリズムによって計算する。

[関連度計算アルゴリズム]

1. $i = 1, Rset(noun) = \{\}$ と初期化する。
2. *i* 番目の文の関連度を、式 (1) で与える。ただし、 $Freq_i(w)$ は、単語 *w* の *i* 番目の文における出現頻度、 α は関連語の重みを表すパラメータ^{*2}とする。

$$Relate_i(noun) = Freq_i(noun) + \alpha \times \sum_{r \in Rset(noun)} Freq_i(r) \quad (1)$$

3. *i* 番目の文に、名詞 *noun* が含まれていれば、その文に含まれているその他の名詞をすべて、 $Rset(noun)$ に追加する。
4. テキストの最後の文に達していれば終了。そうでなければ、 $i = i + 1$ として、2へ。

すなわち、指定された名詞と一文中で共に使われた名詞を、関連語に加えながら、各文の関連度を与えていく。これは、テキストを前から順に読み進める際に、そこまでの内容から、単語間の関係が理解できるかどうかを表す指標として、そこまでに同じ文中で使われた単語同士であれば関係が理解できる、との考えに基づいている。

3.3 各名詞による各文の光量の計算

本節では、各名詞の各文との関連度をもとに、各文を照らす光の量を計算する方法について述べる。

光は、指定された名詞の関連度をもとに、テキストを表示する際の各文の背景色として表 1 の 256 段階の RGB 値^{*3} を与える。すなわち、赤 (R) と緑 (G) と青 (B) の組み合わせによって黄色の光を作成する。また関連度が 0 の時には、背景が真っ黒になり、テキストも黒色で表示される場合、テキストは全く見えな

表 1 各文の関連度に対する背景の RGB 値

関連度	R 値	G 値	B 値
4.0 以上	255	255	55
3.0 以上 4.0 未満	220	220	0
2.0 以上 3.0 未満	190	190	0
1.0 以上 2.0 未満	120	120	0
0.1 以上 1.0 未満	100	100	0
0	0	0	0

くなる。

3.4 名詞集合による各文の光量の計算

名詞集合 *N* による *i* 番目の文の光量は、集合内の各名詞の *i* 番目の文の関連度を合計した式 (2) の関連度 $RELATE_i$ を、表 1 に適用した値とする。

$$RELATE_i = \sum_{noun \in N} Relate_i(noun) \quad (2)$$

すなわち、名詞集合中の 1 つの単語と強く関係している文や、名詞集合中の多くの名詞と関係している文に、大きな光量が与えられる。

3.5 出力

本システムの入出力インターフェースを図 2 に示す。インターフェースの右側に、光と影をつけたい文章テキストを、コピー&ペーストで貼付けて入力を与える。インターフェース左下の「解析」ボタンを押すと本システムの処理が実行され、光と影をつけたテキストを、インターフェース左側に出力する。図 3 に、テキストの可視化例を示す。光と影をつける名詞集合として、タイトルが指定されている場合には、タイトル中の名詞集合を用い、タイトルがないテキストの場合には、テキスト全体によく光を当てる、式 (3) の値が高い名詞 5 つを用いる。ただし式中の *n* は、テキストの文の数とする。

$$text(noun) = \sum_{i=1}^n Relate_i(noun) \quad (3)$$

またインターフェース下部には、光を与える名詞を変更するための、名詞リストとチェックボックスがある。左から、「タイトル中の名詞」「式 (3) の関連度が高い名詞」「出現頻度の高い名詞」「多くの文を光らせる名詞の組合せ」「現在チェックされている名詞」の 5 つのリストを表示している。ユーザはこの名詞リストを用いて、選択する名詞とテキストとの関係を視覚的な光と影によって捉え、各名詞とテキストとの関係の理解に役立てる。

なお、「多くの文を光らせる名詞の組合せ」には、次の Greedy アルゴリズムを用いている。

1. 名詞集合 $C = \{\}$ とする。
2. 名詞集合 *C* によって光が与えられる文 (R 値と G 値が 0 でない文) に比べ、集合 $C \cup \{noun\}$ によって、光が与えられる文の数が最も増える名詞

*1 名詞の中でも、形態素解析により、一般、固有名詞、サ変接続、形容動詞語幹、副詞可、接尾-助数詞に分類される単語を対象とした。

*2 実験的に $\alpha = 0.5$ としている。

*3 RGB 値とディスプレイ上の輝度は非線形の関係にあるため、見たい目の輝度差を元に調整した値となっている。

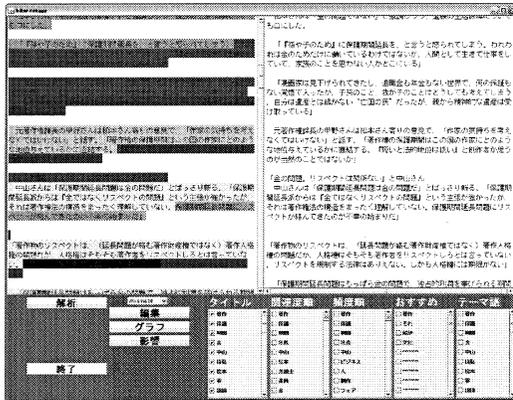


図 2 インタフェースの概観

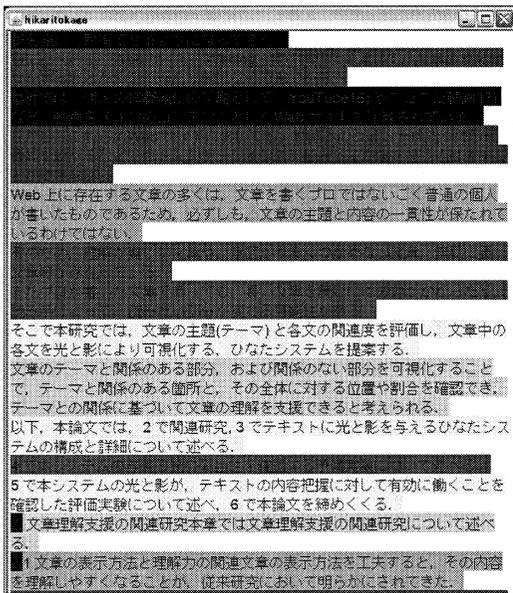


図 3 テキストの可視化例

noun を、集合 C に加える。光が与えられる文が増える名詞がなければ終了。

3. 名詞集合 C によって光が与えられる文の数が、テキストの文の総数の 9 割以上となれば終了。そうでなければ、2へ。

3.6 想定する本システムの使用法

想定する本システムの使用法の主なものとして、文章理解支援、および文章のタイトル作成支援が挙げられる。

3.6.1 文章理解支援

テキスト中の、自分の興味に合う部分だけを素早く探して読みたいユーザを仮定する。

表 2 実験に用いたテキストとその特徴

カテゴリ	一貫性	話題数	話の流れ
ニュース	高い	1つ	あり
ブログ	低い	複数	あり
レビュー ¹⁸⁾	高い	1つ	なし
2ちゃんねる ¹⁹⁾	低い	複数	なし

テキストを入力してシステムに与えると、ユーザは光が当たっている明るい部分だけに着目し、暗い部分を飛ばして読むことができ、内容のすばやい把握が期待できる。またユーザは、再度自分の興味ある名詞を名詞リストから選んで指定することで、自分の興味に関連するテキスト部分を、優先的に探して読むことができる。

また光と影が与えられたテキスト全体をスクロールして見渡すことで、テキストの何割程度に光が当たっているか、またテキスト中のどの部分に特に光が当たっているかを確認して、自分の興味とテキストとの関係を理解することができる。

3.6.2 文章のタイトル作成支援

あるテキストに適切なタイトルを付けたいユーザがいると仮定する。

タイトルは、テキスト全体をよく表す表現にしたいと考えるため、テキスト全体に光を与えられる名詞は参考になると考えられる。すなわち、テキスト全体に光が当たるような名詞の集合を、自分の好みに応じて選んでいくことで、タイトルに含めるべき単語の集合を用意することができるようにと考えられる。

4. 予備実験

本章では、提案するひなたシステムが、指定した名詞と関係のある文に与えられる関連度が、光と影の出力のために妥当な値となっていることを検証した実験について述べる。

4.1 実験内容

表 2 の 4 つのカテゴリの合計 16 テキスト (各ジャンル 4 テキスト) に対して、被験者に各テキストのタイトルと関連のある文を、テキスト中から選んでもらい、システムの出力する光と影が与えられる文と比較する実験を行なった。各カテゴリのテキストの特徴について、ニュースとレビューは明確な 1 つのものについて述べられているので、一貫性が高く話題が 1 つのテキストとした。話の流れの有無は、ニュースとブログは一人の人が書いているのに対して、レビューと 2 ちゃんねるは複数人が書いたテキストとなっていることを元にした。

被験者は情報科学を専攻する大学生、大学院生 28 名で、テキストを前から順に一文ずつ、後戻りしないで読んでもらったときに、全ての文に、タイトルに関係がある、ない、どちらともいえないの 3 択で回答し

表 3 被験者が回答した関連文の数

カテゴリ	関連文の数 (割合)	全文数
ニュース	109(83%)	131
ブログ	40(46%)	87
レビュー	107(68%)	158
2ちゃんねる	298(40%)	746

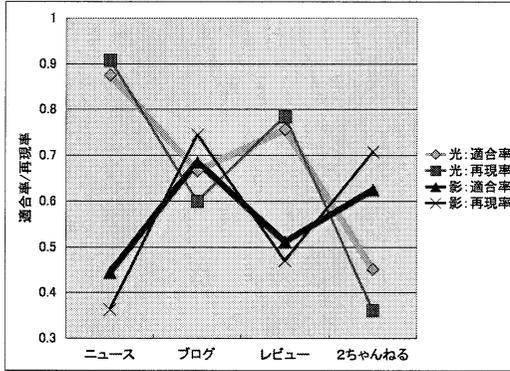


図 4 システムの光と影の適合率と再現率

てもらった。一人の被験者には4つのテキスト(各カテゴリ1テキストずつ)を評価してもらったため、1テキスト当たりの回答数は7名分となった。

4.2 実験結果

過半数の被験者が、タイトルと関係があると回答した関連文の数を表3に、システムの光(タイトルに含まれる名詞集合と関連する文)と影(光以外の文)の適合率と再現率を表4に示す。ただしシステムの光としての出力は、式(2)によって0.1以上の関連度が与えられた文、影としての出力は、その関連度が0となった文とした。

4.3 考察

表4のシステムが出力する光の適合率と再現率について、1つの話題について一貫性がある、ニュース、レビューにおいて高い値となった。これは、本システムが同じ文中に存在する名詞を手がかりとして関連度を与えているため、一貫性があるテキストにおいては似た単語が使われやすく、関連がある文をうまく抽出できたことによると考えられる。

一方、ブログ、2ちゃんねるにおいては、表3において関連する文の割合がともに低くなっており、光の適合率、再現率は、ニュース、レビューほど高い値とはならなかった。特に2ちゃんねるにおいては、複数人で雑談をすることが多く、前の書き込みを反映して議論が発展する傾向が低かったことにより、最も値が低くなったと考えられる。

しかし、ブログ、2ちゃんねるにおいては、表4の影の適合率と再現率が、ニュース、レビューに比べて高い値となった。これは、ニュースやレビューにおい

表 4 文章理解実験に用いたテキストと文数

タイトル	文数
A:衝撃的事実 士農工商はなかった 武士は誰でもなれた	1947
B:普通の電池は充電しても使える.	1342
C:良い電子辞書教えて!	1300
D:ネットから長文が消えたいいくつかの理由	71
E:データは消えない—メモ리카ードや USB メモリに潜む落とし穴	52
F:著作権保護期間は「金の問題」? 中山信弘氏や松本零士氏が議論	164

表 5 各テキストにおける、文の関連度の分布

関連度/TEXT	A	B	C	D	E	F
4.0 以上	4	125	49	1	15	22
3.0 以上 4.0 未満	28	128	82	0	13	15
2.0 以上 3.0 未満	138	199	195	9	12	32
1.0 以上 2.0 未満	376	411	278	28	11	61
0.1 以上 1.0 未満	613	480	697	34	2	35
0	788	994	916	47	4	29

て、タイトルとの関係がやや乏しい前置きや、例示にも、タイトル中の名詞が含まれていたことが原因として挙げられる。逆に、ブログや2ちゃんねるにおいては一文が短く、タイトル中の名詞がない文が積極的に除かれたことによって、値が高くなったと考えられる。

以上より、本システムは、ニュースやレビューのように一貫性の高いテキストについては光を、一貫性が低いテキストについては影の出力を参考にすることができ、テキストの話の一貫性の有無によらず、光と影の出力による一定の効果が期待できることがわかった。

5. 評価実験

本章では、提案するひなたシステムが出力する、光と影の有効性を確認するために行った、文章理解支援、文章タイトル作成支援の2つの実験について述べる。

5.1 文章理解支援実験

5.1.1 実験内容

情報科学を専攻する大学生、大学院生20名に、テキスト中から、テキストのタイトルに関連して重要であると思う文を5文抜き出してもらった。重要な文を抜き出すためには、テキストの内容を理解して要点を捉える必要があると考え、そのためにかかる時間と抜き出された文をもとに評価する。実験には表4のタイトルのテキスト(A,B,Cは2ちゃんねる、D,E,Fは3000から5000字の、あるテーマに関して説明が書かれたコラム)を用いた。制限時間を5分とし、制限時間を超えても5文抜き出せてない場合は、さらに時間を延長して、5文抜き出せるまでの時間を計測した。

各被験者には、以下の提案システムと比較システムとを用いてもらい、6つのテキスト全てについて、指定

表 6 5 文抜き出す時間が 5 分を超えた人数 (10 人中)

テキスト	提案システム	比較システム
A	4	7
B	4	7
C	1	6
D	2	2
E	2	0
F	1	5

表 7 2 人以上が抜き出した文の数

SYSTEM \ TEXT	A	B	C	D	E	F
提案システム	10	8	12	10	12	10
比較システム	11	8	11	9	10	13
共通の文	3	2	0	5	7	9

表 8 関連度 (式 (3))3 以上の抜き出された文の数

テキスト	提案システム	比較システム
A	21	3
B	42	29
C	38	4
D	0	0
E	50	50
F	41	51

されたいずれかのシステムを用いて回答してもらった。

- 提案システム：テキスト全文を表示する。各文にテキストのタイトル中の名詞との関連度 (式 (2)) にもとづいて、光と影を出力する。
- 比較システム：テキスト全体の 20%の文を表示する。提案システムが文に与える、式 (2) の評価値が高い文のみを、出現順に並べて表示する。

5.1.2 実験結果

被験者が 5 文抜き出すまでにかかった時間が 5 分を超えた人数を表 6 に示す。また、2 人以上が抜き出した文の数を表 7 に、関連度 (式 (2))3 以上の抜き出された文の数を表 8 に示す。

5.1.3 考察

表 6 から、A,B,C,F の 4 つのテキストについて、提案システムは比較システムに比べて、5 文抽出を 5 分以内で終わられる人数が多くなった。これは、提案システムにおいては、比較システムの 5 倍の量のテキストを表示しているにもかかわらず、ユーザは光と影を参考にすることで、内容把握ができたためと考えられる。

D,E の 2 つのテキストにおいては、比較と提案、いずれのシステムにおいても、ほぼ 5 分以内で抽出できる結果となった。これは、この両テキストが他のテキストに比べて短く、比較システムにおいてはスクロールなしで 1 画面ですべてのテキストを閲覧できる状態であったためと考えられる。

このことから、画面スクロールが必要な量のテキス



図 5 タイトル作成支援実験に用いたインタフェースの概観
比較システムでは、左上の可視化領域は使用しない

トの内容把握においては、本システムが出力する光と影によって、内容把握が支援できると考えられる。

表 7 において、提示する文の数が多かった提案システムにおいて、抽出された文のばらつきが比較システムと同程度となった。このことから、光と影の出力によって、複数の人が重要と思える客観的に重要な文を、多くの文の中から選んでいることがわかる。2 ちゃんねる以外の D,E,F の説明的テキストにおいて、半数以上は共通の文が抽出されており、流れのあるテキストにおいて、比較システムと質の面で遜色のない内容把握が行なえたと考えられる。

また表 8 から、2 ちゃんねるのテキスト A,B,C においては、提案システムでタイトルとの関連度の高い文が多く選ばれる結果となった。このことから、非常に長く一貫性のないテキストにおいては、重要な文を選ぶための指針がないと、うまく重要な文を選べない可能性が高いことがわかり、本システムの出力する光と影は、そのようなテキストの理解にも役立てられる。

5.2 タイトル作成支援実験

5.2.1 実験内容

情報科学を専攻する大学生、大学院生 20 名に、あるテキストのタイトルに使用すべきだと思う名詞を、5 個以上 10 個以下で選んでもらい、選んだ名詞の出来るだけ多くを使ってテキストの内容を良く表すタイトルを作成してもらった。タイトル作成にかかった時間、選ばれた名詞、作成されたタイトルによって評価を行なった。タイトルを作成するテキストには、表 9 の 3000 字から 5000 字の比較的長い、あるテーマに関して説明が書かれたコラムを用いた。

各被験者には、6 つのテキスト全てについて、以下の提案システムと比較システムのいずれかを用いて、提示される名詞リスト横のチェックボックス (図 5 右下) を使って、タイトルにふさわしい名詞を選んでもらった上で、タイトルを作成して回答してもらった。

表 9 タイトル作成実験に用いたテキストと文数

タイトル	文数
G:医療の「テレビドラマ」は増えてます。でも「報道」が日本医療をダメにしている？	48
H:日本の男はなぜ勝負に弱いのか 五輪野球・サッカーで考えた	78
I:「グレシャムの法則」から見た 基軸通貨ドルの明日	61
J:デフレ下の販売奨励金廃止で 携帯メーカーは敵対的買収の餌食になる	90
K:首都圏のスーパーが電子看板で売上増 ソニーが開拓する新たな「広告メディア」	59
L:取材の方法論を変えたハイビジョンカメラ	115

表 11 タイトル作成までにかかった時間(秒)の中央値

テキスト	提案システム	比較システム
G	449	399
H	486	555
I	506	562
J	710	535
K	559	430
L	588	402

- 提案システム：テキスト全文と、関連度(式(2))が高い順に並べられた名詞リスト、および選択された名詞集合によるテキストの光と影を表示する。
- 比較システム：テキスト全文と、出現頻度順に並べられた名詞リストを表示する。

5.2.2 実験結果

4人以上の被験者がタイトルに使った名詞を表10に、タイトル作成までにかかった時間の中央値を表11に示す。またタイトル作成のために選択された名詞、および実際のタイトルに使われた名詞について、他方のシステムに比べて2人以上多くの被験者が使用した名詞の、テキスト内での出現範囲^{*1}の全テキストに対する割合(被験者平均)を表6に示す。

5.2.3 考察

表10においてタイトルに使われた名詞は、表9の実際のタイトルと比べ、いずれのシステムについても、テキストの内容をおよそ捉えたタイトル付けが行なえていることが確認できる。

表11から、G,J,K,Lのテキストについては、提案システムの方がタイトル付けに時間がかかる結果となった。これは、提案システムにおいては、選択する名詞を変更することにより、光と影の表示が変わるため、さまざまな名詞について、光と影の当たり方を確認していたためと考えられる。

これに関連して表6では、提案システムにおいて、名詞選択の段階でテキストH,I以外、最終的なタイト

*1 テキスト中で最初に出現した文の番号と、最後に出現した文の番号の差

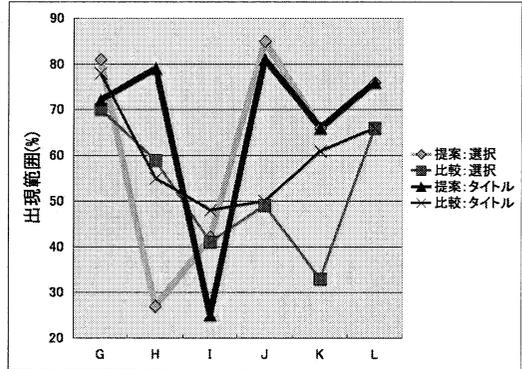


図 6 タイトル作成のために選択された名詞、および実際のタイトルに使われた名詞について、他のシステムに比べて2人以上多くの被験者が使用した名詞の、テキスト内での出現範囲の全テキストに対する割合(被験者平均)

ル中の名詞ではテキストI以外において、テキストの6割以上の範囲で出現する名詞が、比較システムに比べて積極的に用いられていた。このことから、ひなたシステムは、テキスト全体をよく表す名詞を使ったタイトル作成を支援できたと言える。

Iのテキストは「『グレシャムの法則』から見た基軸通貨ドルの明日」という実際のタイトルに対して、「悪貨は良貨を駆逐する」というグレシャムの法則に関連した「悪貨」「良貨」という単語が、テキストの後半になって初めて使われていたため、システム間で出現範囲の差が出なかったと考えられる。しかし、提案システムでは「悪貨」を4人「良貨」を3人が用いており(比較システムでは各1人)、テキスト中の重要な部分を照らす単語の選択を支援できたと考えられる。

6. 結論

本論文では、ユーザが考えるテーマ(名詞集合)に關係するテキスト中の文を、光と影を用いて可視化するひなたシステムを提案した。本システムが文章の理解支援に有効であること、またテキストのタイトルづけに有効であることを、実験により検証した。

今後は提案システムを改良して、文章推敲支援や文章作成支援に役立てていきたいと考えている。

参考文献

- 1) K. Knight, and D. Marcu, Summarization beyond Sentence Extraction: A Probabilistic Approach to Sentence Compression, Artificial Intelligence, Vol.139, No.1, pp.91-107, 2002.
- 2) M. Hotta, Mapping Policy Discourse with CRANE: A Spatial Understanding Support System as a Medium for Community Conflict Resolution, Environment and Planning B:

表 10 4人以上の被験者がタイトルに使った名詞

テキスト		名詞
G	提案	不足, 意志, 深刻, メディア, 産婦人科
	比較	医師, 医療, 不足, メディア
H	提案	勝負, 競争, 日本人
	比較	勝負, 教育, 日本人
I	提案	ユーロ, サンマリノ, 悪貨, 通貨
	比較	価値, ユーロ, サンマリノ, 需要, 通貨
J	提案	携帯, 端末, デフレ, 経済, 産業, 市場
	比較	携帯, 端末, 経済, 構造, 日本,
K	提案	顧客, メディア, 広告, 新た
	比較	業界, 小売
L	提案	カメラ, 機能, 取材, ハイビジョン, ビデオカメラ
	比較	カメラ, ハイビジョン, 取材, ビデオカメラ

- Planing and Design , Vo.27, No.6, pp.801-814, 2000.
- 3) 石岡恒憲, 日本語小論文の論理構成の把握とその図式表現, 人工知能学会論文誌, Vol.23, No.5, pp.303-309, 2008.
 - 4) W. C. Mann. and A. T. Sandra, Rhetorical Structure Theory: Toward a functional theory of text organization, Text, Vol.8, No.3, pp.243-281, 1988.
 - 5) 松村真宏, 加藤優, 大澤幸生, 石塚満, 議論構造の可視化による論点の発見と理解, 知能と情報, Vo.15, No. 5, pp.554-564, 2003.
 - 6) J. Burstein, K. Kukich, S. Wolff, C. Lut, M. Chodorow, L. Braden-harder and M. D. Harris, Automated Scoring Using a Hybrid Feature Identification Technique, in Proc. of the Annual Meeting of the Association of Computational Linguistics, pp.206-210, 1998.
 - 7) 大竹清敬, 岡本大吾, 児玉充, 増山繁, 自由作成要約に対応した新聞記事要約システム YEL-Low, 情報処理学会論文誌「データベース」, Vol.43, No. SIG2(TOD13), pp.37-43, 2002.
 - 8) 関友作, 赤堀侃司, テキストにおける段落表示が内容理解に与える影響, 日本教育学雑誌, Vol.20, No.2, pp.97-108, 1996.
 - 9) 関友作, テキストの内容把握に対する簡条書きとキーワード強調の影響, 日本教育学雑誌, Vol.21, No.2, pp.17-20, 1997.
 - 10) M. S. Smith and E. Vela, Environmental Context-Dependent Memory: A Review and Meta-Analysis. Psychonomic Bulletin & Review, Vol. 8, pp.203-220, 2001.
 - 11) 砂山渡・谷内田正彦, 文章の特徴を表すキーワードを発見して重要文を抽出する展望台システム, 電子情報通信学会論文誌, Vol.J84-D-I, No.2, pp.146-154, 2001.
 - 12) 内田友幸, 田中英彦, 可読性向上を図る対話的文書自動彩色システム, 電子情報通信学会論文誌 D-II, Vol.J80-D-II, No.12, pp.3173-3180, 1997.
 - 13) 魚崎祐子, 野嶋栄一郎, 下線ひき行為が文章理解に及ぼす影響, 日本教育学雑誌, Vol.24, pp.165-170, 2000.
 - 14) D. R. Radev, E. Hovy, and K. McKeown, Introduction to the Special Issue on Summarization, Computational Linguistics, Vol.28, No.4, pp.399-408, 2002.
 - 15) 吉見毅彦, 奥西稔幸, 山路孝浩, 福持陽士, 表題へのつながりに基づく文の重要度評価, 自然言語処理, Vol.6, No.1, pp.43-57, 1999.
 - 16) YouTube:(URL) <http://www.youtube.com/>.
 - 17) ニコニコ動画:(URL) <http://www.nicovideo.jp/>
 - 18) 価格.com: (URL)<http://kakaku.com/>.
 - 19) 2ちゃんねる:(URL)<http://www.2ch.net/>.
 - 20) 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 高岡一馬, 浅原正幸: 形態素解析システム『茶筌』, Version 2.2.9, 使用説明書, (2002).