構造を用いた英語パラグラフ検索機能の評価

An Evaluation of an English Paragraph Retrieval Function Using Structure

國近 秀信 [1*]　　片岡 拓也 [2]

Hidenobu KUNICHIKA[1], and Takuya KATAOKA[2]

[1] 九州工業大学情報工学研究院
[1] Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology
[2] 九州工業大学情報工学部
[2] Computer Science and Systems Engineering, Kyushu Institute of Technology

Abstract: One of the most useful support methods when beginners of English learning write English paragraphs is to refer to appropriately structured paragraphs. In order to support such learners who do not have sufficient knowledge, we have implemented an English paragraph retrieval system based on the structures of paragraphs. This system compares the structure of an English paragraph identified by human with the typical structures of English paragraphs, and calculates the degree of matching in terms of the type and sequence of the components. This paper presents an evaluation of a retrieval function. As the result of the evaluation, it was found that the function can retrieve paragraphs which have desired structures.

1. はじめに

外国語として英語を学ぶ者にとって，論理的で説得力のある英語文章を書くことは困難な作業である．そのような問題の解決法の一つとして，英語の論理展開法に合致した説得力のあるパラグラフを適宜参照することが考えられる．参照するパラグラフは，構造および内容について，学習者が望むパラグラフであることが望ましいと考える．パラグラフライティングに関する書籍には良いパラグラフが掲載されているが，数が限られているためユーザの目的に合致したパラグラフを参照できるとは限らない．また，学習者が書きたい内容に近いパラグラフを得る方法として，WWW 上のパラグラフを検索する方法が考えられるが，ユーザ自身で論理展開法の適切性を判断することが難しいため，一般的なキーワードによる検索では適切な構造のパラグラフを得ることは困難である．これまでに英文の構造に着目した用例文検索システム[1][2]は実現されているが，パラグラフの構造に着目した検索システムは見当たらない．そこで我々は，英語パラグラフの構造を用いた検索システムの実現を目指している．本研究では，WWW より自動収集した英語パラグラフの構造が同定された後を想定し，パラグラフの構造を用いて検索を行う検索システムを実現している[3]．本論文では，本システムにおける検索機能の評価について述べる．

2. パラグラフ検索に必要な機能

本研究では，英語の論理展開法の理解が不十分なユーザが，英語の論理展開法を学習しながらパラグラフを書こうとしている状況を想定している．そのため，英語の論理展開法に則ったパラグラフで，かつ，ユーザが望む種類のパラグラフを検索できるようにする必要がある．また，パラグラフの種類による違いや，構成要素の細かな違いを確認するために本システムを利用することも想定されるため，パラグラフの種類や構成要素を複数指定できる必要がある．さらに，ユーザがアイデア（パラグラフで記述する内容の小片に相当する）を自由に書き出した後で，パラグラフの主題など，方向性を決定するために，実際のパラグラフに書かれている内容やトピックに対する筆者の考えを参考にしたいという状況が考えられる．そのため，ユーザが書きたい内容に近いパラグラフを参照できるようキーワード検索機能

* 連絡先：　九州工業大学情報工学研究院
〒820-8502 福岡県飯塚市川津 680-4
E-mail: kunitika@ai.kyutech.ac.jp

が必要である．最後に，ヒットしたパラグラフの中から，ユーザが必要とするパラグラフを見つけやすくするための機能が必要である．

3. パラグラフ検索システム

本システム[3]の概要を図 1 に示す. 本システムは，パラグラフ展開スキーマ, パラグラフデータベース，パラグラフ検索機能，および，表示機能から成る．パラグラフ収集部では，WWW 上からパラグラフを収集し，それらの構造分析を行う．その後，構造分析済みのパラグラフをパラグラフデータベースへ保存する．パラグラフ検索部では，まず，ユーザが検索したいパラグラフの種類とキーワードなどの条件を指定する．その後，パラグラフデータベース内に保持される構造同定済みのパラグラフと英語パラグラフの典型定な構造を表すパラグラフ展開スキーマとの一致度を算出し，指定されたソート順で出力する．なお，パラグラフ収集部は未完成であるため，現時点では手動でパラグラフの構造を分析し，パラグラフデータベースへ保存している．本章では，各機能について説明する．

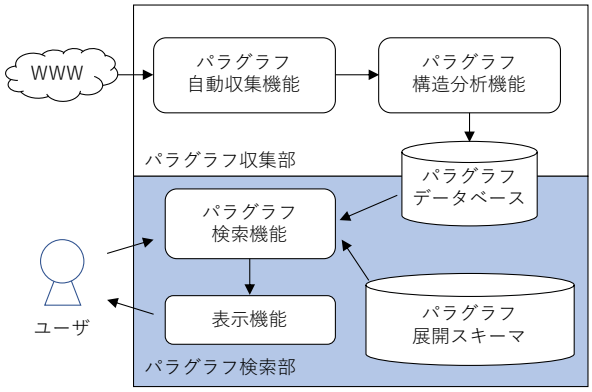

図 1　システムの概要[3]

3.1　パラグラフ展開スキーマ

構造に着目したパラグラフ検索を行うためには，英語の論理展開法に関する知識が必要となる．パラグラフにはその目的・内容によりいくつかの種類があり，それぞれ構成が異なる．我々はこれまでに，パラグラフライティングに関する書籍を基にパラグラフの論理展開法を整理・分類して，人および計算機により解釈可能なパラグラフ展開スキーマを定義した[4]. 本研究で扱うパラグラフの種類は, Listing, Example，Comparison & Contrast (Block Organization, Point-by-Point Organization)，Objective Analysis，Cause and Effect，Opinion and Reason，Definition，Classification，Process and Direction および Personal Description であり，それぞれについてパラグラフ展開スキーマを定義した．各パラグラフ展開スキーマは，典型的な構造を表す Structure，その構成要素の説明文である Explanation，パラグラフを書く際の注意点やコツを表す Tip，頻出語を表す Words and Phrases，および，構造に関する制約を表す Dependence から成る．ここで Structure には，パラグラフの種類ごとに，必要なアイデアの関係・役割，記述順序，数の制約が書かれており，パラグラフの種類による構造の違いが表現される．なお，パラグラフ展開スキーマの適切性については，英語教員により確認済みである．

例として，Listing パラグラフの Structure を図 2 に示す．Listing パラグラフは，重要事項や論点，事例等を並べて説明する場合に使用される．ここで，同図において，四角形は構成要素を表し，構成要素の右肩の数値／記号は，繰り返しの回数を表す（*は 0 回以上，+は 1 回以上を表す）.

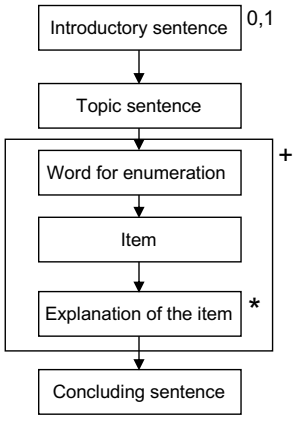

図 2　パラグラフ展開スキーマの Structure の例[3]

3.2　パラグラフデータベース

パラグラフ構造分析機能では，パラグラフの構成要素である句，節，文に対しパラグラフ中での役割を同定し，その役割を表すラベルを付与する．それらの結果は, XML 形式でパラグラフデータベースにて保持される．なお前述の通り，現時点では，人手で構成要素の役割を同定しラベルを付与したものを利用している．

3.3　パラグラフ検索機能

3. 3. 1　構造による検索

ユーザが指定したパラグラフの種類に対応したパ

ラグラフ展開スキーマの Structure を元に，その構造と近い構造を持つパラグラフを検索する機能である．パラグラフ展開スキーマの Structure は典型的な構造であるため，検索対象のパラグラフの構造と完全に一致するとは限らない．つまり，構成要素の種類や数，並びが異なる場合がある．よって本研究では，両者の構造の一致度を求め，検索に利用する．本研究で扱う構造はパラグラフの構成要素とその並びから成ることから，構成要素の種類を用いた一致度，および，構成要素の並びを用いた一致度を求め，それらの平均を最終的な一致度とする．

図 3 に，処理の概要を示す．本システムでは，一致度算出処理の簡単化のため，前処理として，比較対象のパラグラフを分析して数に制約のある要素の個数を同定し（図 3 (a)），そのパラグラフ用にパラグラフ展開スキーマを変形する（図 3 (b)）．例えば図 3 に示すように，Listing のパラグラフ展開スキーマには，数の制約のある要素として，Introductory sentence, Explanation of the item，+記号が付与された要素の組があり，比較対象のパラグラフ中には，それぞれ，0，1，2 個存在することがわかる．その後，その情報を元にパラグラフ展開スキーマを変形し，変形スキーマを生成する．以下，変形スキーマを用いた一致度の計算（図 3 (c)）について述べる．

(1) 構成要素の種類を用いた一致度

比較対象のパラグラフと変形スキーマとを比較し，「比較対象のパラグラフに不足している要素の数」と「変形スキーマに含まれない余分な要素の数」をカウントし．0 から 100 の範囲で一致度を求める．

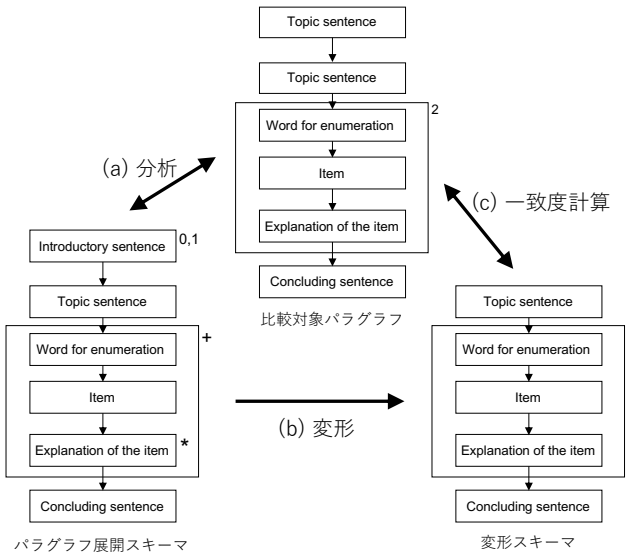

図 3　パラグラフ展開スキーマの変形

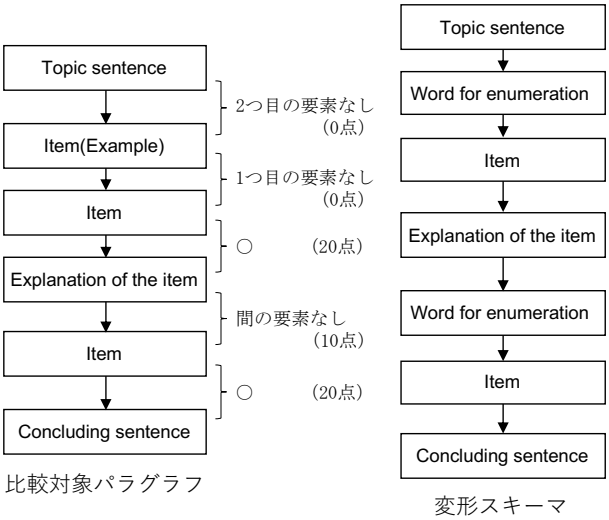

図 4　構成要素の並びを用いた計算例

もし不足および過多の合計が変形スキーマの要素の数を超える場合は 0 とする．

(2) 構成要素の並びを用いた一致度

構成要素の前後関係に着目して一致度を求める機能である．具体的には，まず比較対象のパラグラフより連続する構成要素の組を同定し，その総数に応じて 100 点を按分した値をそれぞれの組に与える．次に要素の組が変形スキーマに存在するか否かを調べる．もし存在する場合は持ち点をそのまま与え，2 つの要素の間に異なる要素が存在する場合やその組が存在しない場合は減点する．最後にこれらの得点の合計を構成要素の並びを用いた一致度とする．

図 4 に例を示す．比較対象のパラグラフには，連続する構成要素の組は 5 個存在するため，各組の点数を 20 点とし，変形スキーマの構成要素の比較を行う．まず，Item(Example)は変形スキーマには存在しないため，これを含む組は 0 点とする．また，Explanation of the item と Item の組については，Word for enumeration が含まれていないため，減点し 10 点とする．最後に各組の点数を合計し，最終的な一致度は 50 となる．

3.3.2 複数種類のパラグラフ検索機能

ユーザがパラグラフの種類ごとの違いを比較できるようにするため，複数のパラグラフの種類を指定して検索する機能である．本機能は，データベース内の各パラグラフについて，指定された種類のパラグラフ展開スキーマとの一致度を求め，その中の最大値の種類のパラグラフとして表示する．

3.3.3　キーワード検索機能

ユーザが書きたい内容に近いパラグラフを参照できるようにするための機能である．本機能は，ユーザが指定した全てのキーワードを含むパラグラフを検索結果として表示する．なお，パラグラフの主題や筆者の考えは Topic sentence に書かれるため，キーワードを Topic sentence のみから探すよう指定することも可能である．

3.4　その他の機能

多くのパラグラフがヒットした場合に検索結果の絞り込みを行うことができるようにするため，パラグラフの検索時には，以下に示す条件を指定するこ

図 5　検索条件の入力画面例[3]

図 6　検索結果の画面例[3]

とが可能である．

(1) 長さ：短い（4 文以下），普通（5～10 文），長い（11 文以上）の 3 種類から長さを選択する．

(2) 構成要素：特定の構成要素が含まれるパラグラフのみを検索対象とする機能である．

(3) 一致度の種類：前述の通り，構造を用いた検索として，構成要素の種類を用いた一致度，および，構成要素の並びを用いた一致度の 2 種類の一致度を用いる．ユーザは，これらのうちのいずれかを指定することができる．

(4) 一致度下限値：一致度が極端に低いパラグラフを表示しないようにするため，必要に応じて，ユーザが一致度の下限値を設定し，検索結果として表示されるパラグラフを制限する．

3.5 表示機能

検索機能でヒットしたパラグラフを指定した一致度の降順で表示する．ただしこの方法については，キーワードを用いた検索や，複数種類のパラグラフの検索の場合は，ユーザが必要とするパラグラフを見つけることが困難になる可能性がある．よって，キーワードの出現回数を元にした表示，および，パラグラフの種類ごとの表示も可能としている．

表示される情報は，パラグラフの種類，パラグラフ，パラグラフの構造，パラグラフ展開スキーマの構成要素との差異，一致度および英文数である．ここで，パラグラフの種類，パラグラフ展開スキーマの構成要素との差異，および，一致度については，各パラグラフ展開スキーマとの一致度の中で最大値となったパラグラフ展開スキーマの情報が表示される．また，パラグラフおよびパラグラフの構造については，両者の構成要素の対応関係を確認しやすくするため，下線により強調表示する．

図 5 に検索条件の入力画面例を示し，図 6 に検索結果の画面例を示す．この例では，Listing と Example の 2 種類のパラグラフが検索条件として入力され，その結果が表示されている．

4. パラグラフ検索機能の評価

パラグラフ検索機能の評価として，構造を用いた検索により，指定した種類のパラグラフが検索可能か否かについて確認した．

4.1 評価方法

パラグラフライティングに関する書籍では，パラグラフの種類ごとにパラグラフ例が掲載されているため，これを利用した．具体的には，パラグラフの種類ごとに 10 個，合計 110 個のパラグラフを取り出し，各パラグラフの構造を人手で分析し，データベースに格納した．続いて，パラグラフ検索機能を用いて，全 11 種類のパラグラフ展開スキーマごとに，データベース中のパラグラフとの一致度を算出した．その後，パラグラフ展開スキーマと同じ種類のパラグラフの一致度の平均点と，異なる種類のパラグラフの一致度の平均点を求めた．また，Dunnett の多重比較により，同じ種類の一致度の方が異なる種類の一致度よりも高い値となっているか否かを確認した．

4.2 結果と考察

表 1 に，Dunnett の多重比較により得られた検定統計量を示す．ここで，各列の見出し部分の番号は，最左列のパラグラフの番号に対応する．表 1 の結果より，全ての種類のパラグラフ展開スキーマについ

表 1　多重比較の結果

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)Listing	—	7.31*	7.25*	7.38*	7.23*	7.09*	7.07*	3.08*	6.90*	7.22*	7.30*
(2)Example	7.41*	—	7.31*	7.41*	7.18*	7.16*	7.14*	7.41*	7.21*	7.27*	7.35*
(3)Comparison&Contrast(Block Organization)	7.29*	7.29*	—	0.79*	7.24*	7.23*	7.06*	7.29*	7.29*	7.06*	7.18*
(4)Comparison&Contrast(Point-by-Point Organization)	7.31*	7.31*	1.09*	—	7.26*	7.25*	7.09*	7.31*	7.31*	7.09*	7.20*
(5)Process&Direction	7.46*	7.40*	7.35*	7.46*	—	7.20*	7.18*	7.46*	7.25*	7.32*	7.39*
(6)Personal Description	7.46*	7.40*	7.35*	7.46*	7.23*	—	7.19*	7.46*	7.26*	7.32*	7.39*
(7)Objective Analysis	7.46*	7.40*	7.36*	7.46*	7.23*	7.21*	—	7.46*	7.26*	7.33*	7.39*
(8)Definition	2.87*	7.21*	7.09*	7.21*	7.17*	7.17*	7.06*	—	7.21*	7.06*	7.14*
(9)Classification	7.00*	7.30*	7.23*	7.37*	7.07*	7.05*	7.02*	7.37*	—	7.19*	7.29*
(10)Opinion&Reason	7.47*	7.47*	7.33*	7.47*	7.43*	7.42*	7.29*	7.47*	7.47*	—	7.38*
(11)Cause&Effect	7.52*	7.52*	7.39*	7.52*	7.48*	7.48*	7.35*	7.52*	7.52*	7.36*	—

*p<.05

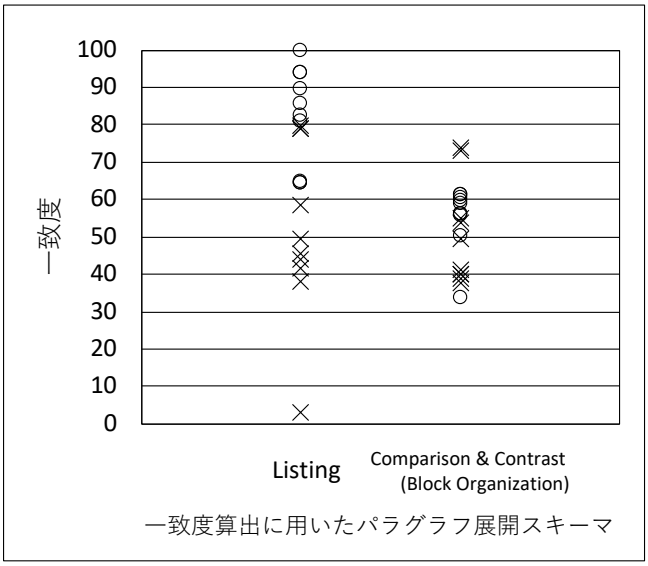

図 7　一致度が近いパラグラフの得点分布

て，同じ種類のパラグラフとの一致度の平均は，異なる種類のパラグラフとの一致度の平均よりも有意に高い値であることがわかる（有意水準 5%）．よって, 構造による検索の際に求められる一致度により，ユーザが指定したパラグラフが検索可能であると判断することができる．

表 1 に示した値のうち, (1) Listing と (8) Definition, および, (3) Comparison & Contrast (Block Organization) と (4) Comparison & Contrast (Point-by-Point Organization)については, 他より低い値となっている．これら 2 つの組み合わせについて，得点分布を図 7 に示す．同図において，○は同種のパラグラフの値を表し，×は異種の値を表す．つまり，Listing のパラグラフ展開スキーマとの一致度を求めた結果については，○は Listing パラグラフとの一致度を表し，×は Definition パラグラフとの一致度を示している．また，Comparison & Contrast (Block Organization)のパラグラフ展開スキーマとの一致度を求めた結果については，○は Comparison & Contrast (Block Organization) パラグラフとの一致度，×は Comparison & Contrast (Point-by-Point Organization)パラグラフとの一致度を示している．

Listing については，Definition のパラグラフが 40 点～50 点近辺に集中している．また，Comparison & Contrast (Block Organization)については，全体的には Comparison & Contrast (Block Organization)のパラグラフが Comparison & Contrast (Point-by-Point Organization)を上回ってはいるが, 両方のパラグラフが 50 点付近に集中しており，80 点を超えるパラグラフは存在しなかった. このように異なる種類のパラグラフを用いた際の得点が高くなってしまった原因は, パラグラフ展開スキーマが似ている点が挙げられる．また，パラグラフ展開スキーマの Structure は，典型的な構造を表しており，さまざまなバリエーションが考えられる. 本調査で利用したパラグラフにおいても, パラグラフ展開スキーマの Structure との差異が確認されたため, その点も影響したと考える.

5. おわりに

本論文では, 英語パラグラフの構造を用いたパラグラフ検索システムにおける検索機能の評価について述べた. 調査の結果, パラグラフ展開スキーマとパラグラフの構造の一致度を用いることにより，指定した種類のパラグラフが検索可能であることが確認された. ただし, パラグラフ展開スキーマは典型的な構造であるため, さまざまなバリエーションが考えられ，一致度により正確に分類することは難しいため，その点に関する配慮が必要となる.

今後は，WWW 等からパラグラフを収集しその構造を分析するパラグラフ収集部を実現する予定である.

謝辞

本研究の一部は, 科学研究費補助金基盤研究 (C)（一般）（No. 19K00763）の援助による.

参考文献

[1] 松原茂樹，加藤芳秀，江川誠二 : 英文作成支援ツールとしての用例文検索システム ESCORT, *情報管理*, Vol. 51，No. 4，pp. 251-259 (2008)

[2] 三好康夫, 越智洋司, 金西計英, 岡本竜, 矢野米雄 : 英作文支援における句構造情報を利用した用例検索ツール，*日本教育工学会論文誌*，Vol.23，No.3，pp.283-294（2003）

[3] 國近秀信, 片岡拓也 : 構造を用いた英語パラグラフ検索システムの実現，*人工知能学会研究会資料 先進的学習科学と工学研究会*，Vol. 90, pp.9-14 (2020)

[4] 國近秀信，斎藤史朗，竹内章 : 英語パラグラフライティングのためのアイデア整理支援, *電子情報通信学会論文誌(D)*, Vol. J102-D, No. 8, pp. 542-552 (2019)