

令和 4 年 6 月 9 日現在

機関番号：17104

研究種目：基盤研究(C) (一般)

研究期間：2019～2021

課題番号：19K00763

研究課題名(和文) 論理展開にもとづく英語パラグラフ検索システム

研究課題名(英文) An English paragraph retrieval system based on logical development

研究代表者

國近 秀信 (KUNICHIKA, Hidenobu)

九州工業大学・大学院情報工学研究院・准教授

研究者番号：70284594

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：英語の初学者が英語文章を書く際に、適切な構造のパラグラフを参照することは、有用な方法の一つである。このような英語の論理展開法を熟知していない学習者を支援するため、本研究では、学習者が欲する英語の論理展開法および内容に合致したパラグラフを提示するシステムを実現した。本システムは、WWWより入手したパラグラフの構造を解析し蓄積する。ユーザは、その情報を元に、特定の論理展開・内容のパラグラフを検索する。

研究成果の学術的意義や社会的意義

本研究では、従来人手に頼るのみであったパラグラフの論理展開に着目した検索について、計算機による自動化を試みた。現時点では制約があるものの、本システムにより、学習者が表現したい構造および内容に近いパラグラフを容易に参照することが可能となった。教育工学の分野では、英語パラグラフライティング支援については、ほとんど支援対象とされていなかったため、本領域の開拓・発展が波及効果として期待できる。

研究成果の概要(英文)：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, an English paragraph retrieval system based on both the logical developments and the contents of paragraphs has been implemented. The system gets English paragraphs from WWW, analyze their structures, and stores them into a database. Users will get paragraphs which have specific logical developments and contents by using the system.

研究分野：教育工学

キーワード：英語パラグラフライティング 検索システム 論理展開 ライティング 英語学習

1. 研究開始当初の背景

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。この主な原因は、英語の論理展開法を熟知していないことだと考えられる。この問題を解決するための一つの方法として、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。一般に文書の内容に関する検索の際には、検索エンジンを用いたキーワードによる検索が行われることが多い。この検索エンジンを用いた方法を英語学習における例文検索に活用することが可能であるが、単語の語法や構文、文章の論理展開の学習を目的とした場合は、不要な検索結果が数多く出力される可能性が高く、特に初学者にとっては利用し難いという問題がある。これまでに文の構造を利用した検索システムが提案されているが、それらのシステムは一文を対象としており、文章の論理展開を考慮した検索には使用できない。

2. 研究の目的

本研究では、WWW上の英語文章を検索し、学習者が欲する英語の論理展開法および内容に合致したパラグラフを提示するシステムの実現を目的とする。本システムは、WWWより入手したパラグラフの構造を解析し蓄積する。ユーザは、その情報を元に、特定の論理展開・内容のパラグラフを検索する。具体的には、以下の点に主眼を置く。

- (a) パラグラフコーパスの実現法の確立
- (b) 論理展開および内容を考慮したパラグラフ検索法の確立
- (c) 検索結果の表示法の確立
- (d) WWW上のパラグラフの自動収集法の確立
- (e) パラグラフの論理展開解析法の確立

3. 研究の方法

(1) パラグラフコーパスの実現法の確立

これまでの研究では、パラグラフの論理展開にもとづく検索ができるようデータベースを実現した。本研究では、さらに学習者の目的に合致した内容のパラグラフを検索できるようパラグラフコーパスを設計・実現する。英語文章の他に、パラグラフの内容に関する検索精度を向上させるための「各英文の構文解析結果」、および、論理展開による検索時に利用する「各英文のパラグラフ中での役割」を保持する。なお、英文の構文解析には一般に公開されている構文解析器を用いる。また、「各英文のパラグラフ中での役割」については、まずは人間がその役割を同定・付与し、後述の(5)にてその自動化に取り組む。

(2) 論理展開および内容を考慮したパラグラフ検索法の確立

学習者の目的に合致した内容および論理展開のパラグラフを検索できるようにするため、内容を表すキーワードおよびその品詞、論理展開を表すパラグラフの種類や特定の構成要素を指定して検索する手法を確立する。パラグラフの種類ごとに用意されたパラグラフ展開スキーマは、英語パラグラフの典型的な論理展開法であるため、その派生型が数多く存在すると考えられる。したがって類似検索の手法についても検討する。

(3) 検索結果の表示法の確立

検索の結果、多くのパラグラフがヒットする可能性があるため、その中から必要なパラグラフへ簡単にたどり着くための表示法を確立する必要がある。先行研究では、パラグラフ展開スキーマとパラグラフとを、構成要素とその並びという観点から比較して類似度を求め、その順番でソートして表示するとともに、学習者が指定した構成要素の色を変更し強調した。本研究では、学習者が入力した内容に関するキーワードも考慮したソート法および強調表示法を確立する。

(4) WWW上のパラグラフの自動収集法の確立

学習者が書きたい内容および論理展開に合致するパラグラフが得られるようにするため、WWW上のHTML文書およびpdfファイルから多くのパラグラフを自動的に収集・蓄積する。その際には、良い英語文章を収集するため、ニュースや学術論文など特定のサイトより収集する。

(5) パラグラフの論理展開解析法の確立

学習者が参照したい論理展開に合致したパラグラフを提示可能とするため、収集したパラグラフの論理展開を解析する必要がある。そのため、まず各英文を構文解析して単語の品詞や係り受けを同定する。その後、キーワードや順序関係などの知識を定義・利用して各文の役割を同定し、タグを付与した上でデータベースへ格納する。この時、タグとして先行研究で定義したパラグラフ展開スキーマの構成要素の名称を利用する。パラグラフ中の文の役割に相当するパラグラフ展開スキーマの構成要素は全部で31種類あり、各構成要素の同定方法を確立する必要がある。この作業は、パラグラフライティング指導書に記載されているパラグラフを詳細に分析した

から進める。

4. 研究成果

(1) パラグラフコーパスの実現法の確立

学習者の目的に合致した内容のパラグラフを、パラグラフの論理展開を利用して検索できるようにパラグラフコーパスを設計・実現した。具体的には、パラグラフ中の英文、および、人手で解析・同定した句／節／文のパラグラフ中での役割を XML 形式で表現することとした。これにより、英語パラグラフの典型的な論理展開法を表すパラグラフ展開スキーマとの比較ができるようになり、また、パラグラフ中の特定の範囲でのキーワード検索ができるようになるため、論理展開を利用した検索、および、学習者の目的に合致した内容のパラグラフ検索が可能となる。

(2) 論理展開および内容を考慮したパラグラフ検索法の確立

学習者の目的に合致した内容および論理展開のパラグラフが検索できるようにするため、内容を表すキーワード、論理展開を表すパラグラフの種類や特定の構成要素を指定して検索する手法を確立した。パラグラフの種類ごとに用意されたパラグラフ展開スキーマは英語パラグラフの典型的な論理展開法であるため、その派生型が数多く存在する。よって、パラグラフ展開スキーマとパラグラフコーパス内のパラグラフとの一致度を求めることとした。また、キーワード検索については、ユーザが必要とするパラグラフを検索できるようにするため、パラグラフの構成要素を指定した検索ができるよう実現した。

さらに、将来の機能拡張に向け、構造比較機能を実現した。ユーザが欲するパラグラフを検索できるようにするためには、ユーザが自由にパラグラフの構造を指定できるようにすることが望ましい。そのような機能を実現できるようにするため、下位機能として構造比較機能を実現した。本機能は、ユーザが入力したパラグラフの構造と、パラグラフデータベース内の各パラグラフの構造とを比較し、条件を満たすパラグラフを出力する。

(3) 検索結果の表示法の確立

検索の結果、多くのパラグラフがヒットする可能性があるため、その中から必要なパラグラフへ簡単にたどり着くための手法を確立した。本システムは、英語パラグラフの典型的な論理展開法との一致度によるソートだけではなく、ユーザの要求により、パラグラフの種類ごとのソート、および、キーワードの出現回数によるソートも行うよう実現した。

(4) WWW 上のパラグラフの自動収集法の確立

学習者が書きたい内容および論理展開に合致するパラグラフが得られるようにするため、WWW 上の HTML 文書からパラグラフを自動的に収集・蓄積する手法を確立した。パラグラフの英文や構造の質はさまざまであるため、本研究では、あらかじめ指定したサイトよりパラグラフを収集することとした。本システムは、特定のサイトにアクセスし、あらかじめ用意した「サイト固有の情報」をもとにパラグラフを取り出す。その後、日時や著者名など、不要な要素を除去する。続いて、論理展開解析機能の利用を想定した上で、英語パラグラフの典型的な論理展開法との一致度を求め、特定の種類の論理展開法のみ一致度が高かったパラグラフをデータベースへ格納するよう実現した。

(5) パラグラフの論理展開解析法の確立

学習者が参照したい論理展開に合致したパラグラフを提示可能とするためには、収集したパラグラフの論理展開を解析する必要がある。そのため、まず各英文を構文解析して単語の品詞や係り受けを同定し、その後、キーワードや順序関係などの知識を定義・利用して各文の役割を同定し、タグを付与するよう実現した。この時、タグとして先行研究で定義したパラグラフ展開スキーマの構成要素の名称を利用した。

パラグラフ中の文の役割に相当するパラグラフ展開スキーマの構成要素は全部で 31 種類あり、各構成要素の同定方法を確立する必要がある。本研究では、パラグラフライティングの指導書に書かれている頻出語句や書き方のコツ、パラグラフ例等を調査し、役割同定のためのルールを抽出し実装した。より具体的には、頻出語句、前後の語・句・文との関係、文の位置、パラグラフの種類を利用した。例えば、「意見と理由パラグラフ」の Topic sentence については、「1 から 3 文目までに当該パラグラフの頻出語句を一つ以上含む場合、最初に出現する頻出語句を含む文を Topic sentence とする」という方法で同定するよう実現した。

評価として、11 種類のパラグラフについて 10 個ずつ、計 110 個のパラグラフを用意し、人手による役割同定との比較を行った。その結果、構成要素ごとの再現率と適合率の平均は、それぞれ 26.2%、42.9%であった。ユーザへの誤った情報提供を避けるため、本研究では適合率を重視する必要があるものの、現状では低い適合率となっているため、今後の改良が必要である。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 國近秀信, 片岡拓也
2. 発表標題 構造を用いた英語パラグラフ検索システムの実現
3. 学会等名 人工知能学会, 先進的学習科学と工学研究会
4. 発表年 2020年

1. 発表者名 國近 秀信, 片岡 拓也
2. 発表標題 構造を用いた英語パラグラフ検索機能の評価
3. 学会等名 人工知能学会研究会, 先進的学習科学と工学研究会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------