

分類パラグラフを対象としたトピックセンテンス診断機能の開発

Development of a Diagnostic Function of the Topic Sentence for a Classification Paragraph

石川 寛大^{*1}, 國近 秀信^{*2}

Kandai ISHIKAWA^{*1}, Hidenobu KUNICHIKA^{*2}

^{*1}九州工業大学大学院情報工学府

^{*1}Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology

^{*2}九州工業大学大学院情報工学研究院

^{*2}Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology

Email: k.ishikawa@minnie.ai.kyutech.ac.jp

あらまし: 説得力のあるパラグラフを書くにあたって, トピックセンテンスはパラグラフのテーマを形作る重要な構成要素である. トピックセンテンスには複数の条件があり, それらとの整合性をとったうえでトピックセンテンスを書くのは英語の初学者にとって難しい作業である. そこで本研究では英語の初学者の学習支援の一環として, 複数の条件を基に学生が入力したトピックセンテンスがパラグラフに適したものであるかを診断するシステムの実現を目的とする. 本システムでは, 入力であるトピックセンテンスを分析し, トピックセンテンスの条件に適したものであるか否かを診断する.

キーワード: 英語学習支援, パラグラフライティング, トピックセンテンス, 分類パラグラフ

1. はじめに

英語で説得力のある意見や主張を表現するための文章作成の手法の一つとして, パラグラフライティングがある. パラグラフにはトピックセンテンスやその指示文など複数の構成要素があり, それらすべてが一貫していることがパラグラフライティングにおいては重要である. パラグラフ中でトピックセンテンスは, 著者の主張を明確に表しパラグラフ全体の方向性を示す重要な役割を持つ構成要素である. トピックセンテンスを書くにあたってその表現にはパラグラフ内での機能を満たすために必要な条件が存在する. また, パラグラフの論理展開法には列挙, 例示, 分類, 叙述などの十種類程度の展開法が存在しており, それらの展開法に応じたトピックセンテンスの条件も存在する. 適切にパラグラフを展開するにはそれらの複数の条件を理解する必要があるが, 英語初学者は知識が不十分である場合が多い. そこで我々は, パラグラフライティングの知識が十分ではない初学者を対象とした英語パラグラフライティング作成支援システムの一つの機能として, トピックセンテンス診断機能の実現を目指している.

本研究では, パラグラフの論理展開法のうち類似した性質や特徴によってトピックに相当する物事をグループに分け, それらを適切に配列し説明する分類パラグラフ⁽²⁾を対象としたトピックセンテンスの診断機能を実装した.

2. 分類パラグラフのトピックセンテンス

トピックセンテンスは, パラグラフ全体が何について書かれたものであるかを示すトピックと, トピックに対しての筆者の主張や意見を表すアサーションの二つで構成される. 本研究では, 分類パラグラフのトピックセンテンスについて, 主にアサーションに関する条件をパラグラフライティングの書籍をもとに調査し, 3種類の条件に整理した. また, 分類パラグラフのトピックセンテンスの形式を2種類に分類した. 3種類の条件はトピックセンテンスの形式によってそれぞれ適用するか否かが異なる. 以下, トピックセンテンスの形式, アサーションの条件について述べる.

2.1 トピックセンテンスの形式

(形式1) トピックを主語に置く形式

“Popular music can be divided into three types.”のように, トピックに相当する単語を主語とし, それが分類されていることを示す動詞が続く形式である.

(形式2) 虚辞を主語に置く形式

“There are three kind of skiing.”のように, Thereなどの虚辞を主語として, それにbe動詞が続く形式である.

2.2 アサーションの条件

(条件1) 分類を意味する動詞を用いていること
分類パラグラフのトピックセンテンスでは, トピックで示された物事がいくつかのグループに分類されるということを示す必要がある. 形式1において, アサーション内でトピックをその動作主に持つ動詞や受動態の動詞は, 分類や構成を意味するものである必要がある. この分

類や構成を意味する動詞には”be divided into”などの頻出語彙⁽³⁾が存在する。形式2ではbe動詞を用いるためこの条件は適用されない。

(条件2) 分類の基準が示されていること

トピックセンテンス内では、トピックに相当する物事を一つの基準をもとに分類する必要がある。文中で分類の基準を明示する必要がある。この分類の基準に相当する単語には type, kind などの頻出語彙⁽²⁾が存在する。

(条件3) 分類の個数が示されていること

パラグラフ内でトピックを複数のグループに分類する際、それらのグループはすべて文中で内容が取り上げられる。そのため、グループの個数は示されている必要がある。この個数の表現方法には、”divided into three types”など個数とその分類の基準が示されているものと、”classified as A and B”など各分類を列挙するものの2種類がある。

3. トピックセンテンスの診断機能

診断機能では、ユーザが入力したアイデアについて構文解析器 CoreNLP⁽¹⁾を適用し、各形態素同士の文法的な関係性を解析する。その結果からアサーションの条件に合致する単語があるか否かを判定しトピックセンテンスが各条件を満たしているか否かを診断する。CoreNLPの構文解析結果の例を図1に示す。入力の形式は、トピックセンテンス1文である。

入力	
“Popular music can be divided into three types.”	
CoreNLPの解析結果	
root(ROOT-0, divided-4)	一文のroot
amod(music-2, Popular-1)	形容詞的修飾詞
nsubjpass(divided-5, music-2)	名目上の主語
aux(divided-5, can-3)	受動態の名目上の主語
auxpass(divided-5, be-4)	受動態の補助助動詞
num(types-8, three-7)	数値修飾詞
prep_into(divided-5, types-8)	前置詞句修飾詞(into)

図1 CoreNLPの構文解析結果の例

トピックセンテンスの診断では、アサーションの形式の判定、各条件の一致の判定の順に判定処理を行う。まず、アサーションの形式の判定では構文解析結果のうち、文のrootである動詞がBe動詞であるか否かと主語が虚辞であるか否かを基に判定する。次に、トピックセンテンスの各条件の判定について、条件1については文の主語に対応する動詞を抜き出し、それが分類を意味する動詞であるか否かを判定する。この判定については実際の分類パラグラフに用いられる頻出の動詞⁽²⁾⁽³⁾⁽⁴⁾⁽⁵⁾と一致するか否かを基に判定する。また、条件1は形式1のアサーションにのみ適応する。条件2については、動詞の目的語や虚辞を主語に持つ単語を分類の基準として抜き出し、それらが実際の分類パラグラフに用いられる頻出単語⁽¹⁾⁽²⁾⁽³⁾⁽⁴⁾と一致するか否かを基に判定する。条

件3は、分類の基準に相当する単語を修飾する数詞修飾語があるか否か、また分類の基準に相当する単語が”A and B”などの列挙形式で記述されているか否かを基に判定する。以上の各条件の判定を基に診断結果を出力する。

出力結果では、入力された文がすべての条件を満たしている場合は、トピックセンテンスの形式の判定結果や分類の基準や個数などの重要な単語を抽出し出力する。入力文が条件を満たしていない場合は、満たしていない条件を指摘する文章を列挙形式で出力する。入力された文が条件を満たしている場合、および、誤りがある場合の出力結果をそれぞれ図2、図3に示す。図2の診断結果では、トピックセンテンスの形式の判定結果や、文章から抽出した分類を示す動詞、分類の基準、分類の個数などの重要な単語を明示している。図3の診断結果では、入力された文が満たしていない条件2、条件3についての誤りを指摘する文章を明示している。

[入力]	
Assertion :Popular music can be divided into three types.	
[判定結果]	
*入力されたトピックセンテンスは適切です。 トピックセンテンスの形式:トピックを主語に置く形式	
分類を示す動詞:	divide
分類の基準:	type
分類の個数:	three

図2 文が条件を満たしている場合の出力結果

[入力]	
Assertion :There are many music in the world.	
[判定結果]	
*入力されたトピックセンテンスはいくつかの条件を逸脱しています。 分類の個数が示されていません 分類の基準を表す単語が music となっており適切ではありません	

図3 誤りがある場合の出力結果

4. おわりに

本研究では、分類パラグラフのトピックセンテンスの診断機能を実現した。今後は、本診断機能の有用性に関する評価や、列挙や例示などの分類パラグラフ以外のパラグラフについてトピックセンテンスの診断機能を実装する予定である。

参考文献

- (1) Stanford NLP Group, CoreNLP, <https://stanfordnlp.github.io/CoreNLP/>(参照 2023.02.01)
- (2) Hajime Iwamoto, Hiroyo Yoshida, “パラグラフ・ライティングからアカデミック・ライティングまで”, Shohakusha (2003)
- (3) Yoshihito Sugita, Richard R. Caraker, ”Primary Course on Paragraph Writing”, Sibido Co. (2007)
- (4) C. Jeriel Howard, Richard Francis Tracz, “Writing Effective Paragraphs”, Winthrop Publishers Inc. (1976)
- (5) Kelly Kennedy-Isem, Osamu Takeuchi, Kaori Nishi, Tomoko Yabukoshi, “Keys to Better Paragraph Writing”, Shohakusha (2005)