

平成 30 年 9 月 19 日現在

機関番号：32706

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00454

研究課題名(和文) 物語構造分析に基づく専門文書の統合的読解技術の向上に関する研究

研究課題名(英文) A Study of Improving Reading Skills for Research Articles based on the rhetorical structure

研究代表者

内山 清子 (UCHIYAMA, KIYOKO)

湘南工科大学・工学部・准教授

研究者番号：20458970

交付決定額(研究期間全体)：(直接経費) 2,700,000円

研究成果の概要(和文)：研究論文の内容を理解するための必須情報を調べるために、研究者や学部生などの間における読みの違いを実験した。どの章で時間をかけて読んでいるのか、ページの推移、専門用語をどの程度調べながら読んでいたのかを分析した。結果として初心者は比較的平均的に時間をかけて読んでいたが、専門家は最初の抄録、実験部分、最後の考察を効率的に読んでいることがわかった。この結果から、抄録に出てくる重要な部分の鍵となる表現を見つけることができれば、より初心者が読みやすくなるのではないかと考え、研究論文でよく使われる表現を抽出した。分野を固定して出現頻度の高い単語や表現をある程度抽出することができた。

研究成果の概要(英文)：We investigate essential information for understanding contents of Research Articles (RAs) and reading behavior between researchers as experts and undergraduate students as novices. We implemented a system for taking logs of reading activities. Experiments were carried out to clarify the reading behavior of experts and novices using the system. The results are given that researchers tend to take a lot of time to read the first page and last page to grasp the outline of the study.

To support novices in reading and understanding the contents of RAs, it is needed to indicate the important understanding points for effective reading. Author's intentions or comments based on previous model identify more detailed structure. We tried to analyze rhetorical structure of RAs and extract typical expressions in Japanese. To integrate establishing expressions might help novices construct the background knowledge of the target field by categorizing each point (important problem, method etc.)

研究分野：自然言語処理

キーワード：専門用語 読解分析 修辞構造

1. 研究開始当初の背景

現在、大学生はインターネットを利用したレポート作成が主流となっており、ネット上の他人の文章を自分の意見として軽い気持ちでレポートや論文にまとめてしまう剽窃行為が蔓延している。自分の考えを自分の言葉で表現するためには、専門文書における特有の言語表現や文章間の関係を読み解くための言語的知識や、自分でその知識を活用できる技術を習得する必要がある。更に大学では、事前知識がない状態から論文などの専門文書の読解による専門知識獲得が求められている。つまり、大学の初年度教育における小論文からレポートへ、専門教育におけるレポートから論文への高次の読解による知識獲得のための枠組みが重要となる。

そこで、事前知識が欠如している熟知度の低い内容の文書において、文や段落の間に存在する論理的な関係(文章構造)や、文章全体の論理展開(論理構造)を把握するトップダウン処理の具体的なステップを明らかにする必要がある。更に、大学で学ぶべき専門分野知識を効率的に獲得した上で統合的に文書内容の理解を深める読解技術が求められている。

2. 研究の目的

本研究は、以下の3点を中心として、読解プロセスにおける基礎的分析に基づきレポートや論文の読解・執筆する時に必要とされる文章・論理構造と専門分野に関する知識を活用した読解技術モデルの構築を目指す。

まず、(1)専門文書を読む時に必要な言語・専門知識についてweb実験に基づいた読解プロセス分析を行う。対象データとなる白書、レポート、論文等を収集し、専門文書のコーパスを作成する。次に、(2)物語構造分析を用いることで専門文書における詳細な文章・論理構造や専門用語の機能を分析する。最後に、専門文書内で(3)各文章・論理構造における特徴的な言語表現の類型化と専門分野知識の体系化に基づいた専門文書における統合的な読解技術モデルの構築を実施する。

3. 研究の方法

初年度には分析対象となる文書収集と読解行動の分析を行うために(1)専門文書読解のための言語データの蓄積と整理を実施する。次年度は、読解技術に必要なトップダウン処理の情報として(2)物語分析手法に基づいた専門用語の機能分析と専門用語間関係の抽出、ボトムアップ処理の情報として(3)文章・論理構造の要素と言語的特徴の抽出を行う。最終年度はトップダウン処理とボトムアップ処理の情報を統合するために(4)言語的特徴の類型化・専門分野知

識の体系化とモデルを構築し、読解実験を実施することで本研究の有効性を示して行く。

平成27年度では、大学生に必要とされる高次の読解において、読解に必要な知識の分析と分析対象となる専門文書の収集と整理を実施する。まず読解に必要な知識の分析については、初心者と専門家に専門用語の重要度と論文内容の理解度を判定する読解実験をすでにWeb上のシステムで行っており、そのデータに基づいて更に詳しい分析を行う。

初心者と専門家における専門用語抽出、文書・論理構造における読解時間の違いなどに加え、ページ遷移を分析することにより、論文を理解しようとする時にどの文章・論理構造に戻るのか、あるいはどの部分を何度も読み返すのかなどを調べる。これにより、初心者が繰り返し読む部分のある程度特定することが可能となる。これは、次年度以降に分析する文章・論理構造と対応させて、初心者が理解しやすい・しにくい構造などの分析に活用することができる。

次に分析対象となる専門文書データの収集・コーパス化については、すでに収集・蓄積している論文データに加え、専門的な内容を一般化・汎用的に記述している読解用の資料および分析対象の手本となる文書として、各省庁で公開されている「白書」をデータとして追加する。「白書」データは日本語書き言葉均衡コーパス(BCCWJ)の中に1976年から2005年まで1,500件分収録されており、このデータは文書構造や形態素情報などの注釈が付与されているため、言語分析にそのまま利用することが可能である。更に、見本となる正例だけでなく、実際に大学生が課題に対して提出したレポートをコーパス化してデータベースに登録する枠組みも構築する。

平成28年度は、読解技術に必要なボトムアップ処理の情報として(2)物語分析手法に基づいた専門用語の機能分析と言語的特徴の抽出、トップダウン処理の情報として(3)文章・論理構造の要素と言語的特徴の抽出を行う。実際は(2)(3)の二つの段階は区別するのではなく並行して研究を進めて行く。

まず、物語構造分析手法の基礎理論を調査し、専門文書に適用できる部分との切り分けを行う。物語構造分析では、登場人物を中心にすえて文や各登場人物間の「役割・機能」を分析し、物語の変遷について文の機能ごとに「シーケンス分析」するなど、物語を類型化するための様々な手法が行われてきた。本研究では、その手法を整理し、専門文書に適用していく。特に専門文書の中には様々な専門用語が出現するが、その専門用語は物語における主人公的な役割を果たしているのか(文の主題)、あるいは贈与者として主人公に付加価値をつける役割を果たしているのか(主題の手法)など、用語自身にも異なる役割があると考えられ、ある分野においては

似たような役割を持つ専門用語が存在していると予測できる。そこで、専門用語が文書の中でどのような機能を果たすのか、どのような目的で使われるのかを分析していく。更に専門用語同士の関係を特定できる言語表現を抽出し特徴を分析する。

次に(3)の専門文書に内在する文章・論理構造の要素と言語的特徴の抽出を行う。これまで行った研究では、学術論文に含まれる「背景、目的、動機、手法、研究比較、データ」などの論理構造の分析を行ってきたが、今回は更に細かい単位として文レベルで分析を行うことによって、文・段落間の文章構造について要素を抽出し整理していく。それと同時に、たとえば、文章・論理構造で「背景」を説明している文章で典型的に用いられる動詞「～が求められている」などに着目して、論理構造を特徴づけている語(接続詞、名詞、動詞など)だけでなく、言語表現も抽出する。言語表現を文章・論理構造ごとに以下の図3のように整理することによりトップダウン処理に必要とされる言語知識を収集しデータベース化する。

平成29年度は、前年度にまとめた専門文書における文章・論理構造の要素における言語表現を類型化することと、専門用語の機能と専門用語同士の関係を表現した分野知識体系を構築していく。具体的には、専門分野毎にまとめた専門用語がその分野においてどのような「役割・機能」を果たしているのかを図4のように図式化することによって分野知識を表現していく。

専門文書の読解において、文書の文章・論理構造の把握と専門用語の機能に着目しながら事前知識の不足を補っていく方法と、事前に分野知識を図式化したものを例示するだけの方法によって、内容を理解する読解実験を行う。事前知識がなくても、言語的特徴や論理展開を予測しながら内容を理解できることを実証する。さらに、専門分野知識を専門用語の定義文だけを与えて内容を理解させる方法と図式化した分野知識を与える方法による理解度の違いなども実験していく。最終的には、言語的手がかりと、専門分野の知識の両方を使ったモデルを用いて統合的な読解技術による効率的な内容理解を目指す。

4. 研究成果

2015年度は、専門分野(情報処理分野と自然言語処理分野)の論文を学部学生や、分野外の学生、研究者に読んでもらい、専門用語について(1)必須である、(2)知っておいた方がよい、(3)後で調べる、(4)Webで調べる、の4つの選択肢を用意し、抽出をしてもらう実験を行った。

この実験では、初心者が専門用語の意味を調べる時に適切に定義などを表示できるようなシステムを開発することを目的として実施した。実験は英語と日本語の論文を用意

し、ページ毎に読む時間、専門用語として重要であるかどうかの判定、ページの推移のログを取り、その結果を分析した。分析結果は、初心者や分野外の学生や研究者にとって、専門用語であるかどうかの判断も難しかったという結果が出た。つまり、初心者は専門用語の抽出ができないため、理解をする以前の問題であることがわかった。

一方で、研究者は分野外であっても比較的多くの専門用語を抽出することができていた。また、初心者はまんべんなく論文全体を読んでいたのに対して、研究者は最初の2ページと最後のページに時間をかけて読む傾向が出ていた。論文の概要を把握するためには、最初や最後を効率的に読む技術が研究者は経験的に培われているのだと考えられる。

2016年度は、物語構造分析手法の基礎理論を調査し、専門文書に適用できる部分との切り分けを行うことを目標にした。まず、物語構造分析では、登場人物を中心にすえて文や各登場人物間の「役割・機能」を分析するなどの手法があり、その手法を論文以外の文書で実施し、文書全体に必須の要素として何が必要かを分析した。

物語の場合は「行動」、「イベント」、「感情」の三つの機能にそれぞれ詳細な要素を分類することができた。この手法を専門文書に適用すると「行動」が研究やレポートにおける「実験した」「分析した」などの具体的な動詞で表現され、「イベント」は専門用語で表現される手法や理論、「感情」は著者の意図を反映した叙述表現と考えられる。

そこで、まず論文やレポートにおける「はじめに」の部分について分析を行った。分析では、「はじめに」の文章の一文が何を示すのかを英語論文のGenre AnalysisやScience Research Writingを参考にして分類した。各文の役割として「目的」「背景」「既存の手法」「問題点」「オリジナリティ」などに分けることができ、それぞれの文において使用される言語表現を分析した。

その結果、「目的」や「背景」では、機能として「行動」に該当する論文やレポートであれば一般的に用いられる表現が多く用いられ、「既存の手法」「問題点」「オリジナリティ」では「イベント」に相当する専門用語や特定分野において用いられる表現などが含まれることがわかった。著者の意図を反映した「感情」表現はどの箇所にも出現する可能性があるが、形容動詞、助動詞、複合辞で表現されているため、識別が他の表現よりも容易であった。

2017年度は、昨年と同様に読解プロセスにおける基礎的分析に基づきレポートや論文の読解・執筆する時に必要とされる文章・論理構造と専門分野に関する知識を活用した読解技術モデルの構築を目指し、物語構造分析を用いることで専門文書における詳細な文章・論理構造や専門用語の機能を分析した。

物語の構造分析というのは、昔のおとぎ話

や映画などのストーリーを分類するために用いられる理論であるが、その理論を論文に適用できないかと考えていた。論文に適用する前に、まず漫画のストーリーについて構造分析を行った。恋愛漫画に焦点を当てて、典型的なストーリーパターンを分析し、抽出を行った。この成果については、人工知能学会で発表を行った。

次に、論文の構造への適応を試みた。論文の構造というのは、一般的に IMRAD (Introduction, Method, Related Works, Discussion)の構造で適用できるが、論文の形式的な構造だけでなく、どのようにしてその論文の特徴や長所を主張するのか、著者の意図を反映するのか、といった点に着目し、より詳細な分析を試みることができるのではないかと考えている。

これは CARS モデル(Creating a research space” というモデルが現存しており、英語で Move や Move における各ステップについて述べられている。これは英語でしか記述がないため、日本語の論文に適用できるものかどうかについて分析を行った。

5. 主な発表論文等 (研究代表者は下線)

〔雑誌論文〕(計 0 件)

〔学会発表〕(計 6 件)

Kiyoko Uchiyama, An Analysis of Expression Patterns for Establishing Research Significance, the 9th International Conference on Information, Process, and Knowledge Management (eKNOW2017), 2017.

Kiyoko Uchiyama, A Study for Analyzing the Function of Introductory Terms and Collocation in Academic Papers, Association for Reading and Writing in Asia, 2017.

Kiyoko Uchiyama, A Study of Analyzing Comprehensive Reading Behavior for Research Activities based on a Web-based Experiment, Proceedings of the Pacific Association for Computational Linguistics (PACLING 2015), pp.132-136, 2015

内山清子・石川諒, 恋愛漫画における物語構造分析とその応用, 人工知能学会第 30 回全国大会, p1-2, 2017.

内山清子, レポートや論文に含まれる基礎語彙の分類に関する一考察, 人工知能学会第 29 回全国大会, 2K1-4, p1-2, 2015.

山崎誠・柏野和佳子、内山清子、他, 『分

類語彙表増補改訂版』へのアノテーション -基本義の決定-, 計量国語学会第 58 回大会, 2015.

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
国内外の別:

取得状況(計 件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
取得年月日:
国内外の別:

〔その他〕
ホームページ等

6. 研究組織

(1) 研究代表者

内山 清子 (UCHIYAMA, Kiyoko)
湘南工科大学・工学部・教授
研究者番号: 20458970

〔その他の研究協力者〕

相澤彰子 (AIZAWA, Akiko)
国立情報学研究所・教授
研究者番号: 90222447