

氏名	鄧 新 宇
学位(専攻分野)	博 士 (情 報 学)
学位記番号	情 博 第 294 号
学位授与の日付	平 成 20 年 3 月 24 日
学位授与の要件	学 位 規 則 第 4 条 第 1 項 該 当
研究科・専攻	情 報 学 研 究 科 知 能 情 報 学 専 攻
学位論文題目	An English Text Generation System for Intermediate Non-native Speakers based on Corpus Analysis (コーパス分析に基づいた中級ノンネイティブ向け英語文章生成システム)
論文調査委員	(主 査) 教 授 松 山 隆 司 教 授 黒 橋 禎 夫 教 授 河 原 達 也

論 文 内 容 の 要 旨

本論文は、日本や中国の高等学校において使われている英語教科書に載せられた英文のコーパスを構築し、文章中の接続詞や句読点の配置、順序に関する特徴分析に基づいて、英語中級ノンネイティブ（英語を母国語としないが英語能力が中級レベルの人）にとって理解しやすい構造を持った英語文章を生成するシステムについて述べたもので、6章から構成されている。

第1章は序論で、英語が国際標準語として普及するにつれ、世界規模で見た場合、ノンネイティブの人口がネイティブの倍ほどにも達し、ノンネイティブにとって理解しやすい英語文章を生成することの重要性が高まっていることを指摘するとともに、文章の読み易さを評価する手法に関して行われた従来研究を概観し、本研究の意義と関連研究を紹介している。また、ノンネイティブ向けの英語文章生成システムとして開発したSILKの全体的構成を示し、本論文では文章の修辞構造を表す木構造から、接続詞や句読点で文が結ばれた文章を生成するという、文章構造生成レベルに焦点を当てて研究を行ったことを述べている。

第2章は、ノンネイティブ向け英語文章として収集した40万語からなるテキストに対して、修辞構造を付加するためのアノテーション法を述べ、その精度評価を行った後、出現頻度の高い6種類の接続詞および句読点の配置を規定すると考えられる特徴として、文章の全体的構造に係る8特徴、埋め込み構造に係る10特徴を定め、これらの特徴の有効性を決定木学習システムC4.5およびサポートベクトルマシンを使った学習、分類実験によって評価している。

第3章では、6種類の接続詞および句読点のそれぞれに対して、修辞構造が与えられた場合、どのような特徴によってその配置が定められるのかをコーパスに基づいて分析し、各接続詞、句読点ごとに文章中での配置の適切性、すなわちノンネイティブにとっての文章の読み易さを表すランク付けルールを導出している。

第4章では、3章で求めた配置の適切性を示すランク付けルールに基づいて設計された評価関数を遺伝子の適合度とする遺伝的アルゴリズムを開発し、与えられた修辞構造からノンネイティブが理解しやすい文章を生成するシステムについて述べている。

第5章では、4章のシステムで生成された文章の読み易さを、システムが生成した文章と元のコーパス中の文章との一致性を評価する方法および、システムが生成した文章が人間にとって読み易いか否かを主観的アンケート調査によって評価する方法の2つの方法で評価し、いずれの評価法においても、システムが生成した文章がノンネイティブにとって読み易いものとなっていることを実験的に明らかにしている。

第6章は、結論であり、本研究のまとめと今後の展望を述べている。

論 文 審 査 の 結 果 の 要 旨

本論文は、日本や中国の高等学校における英語教科書に載せられた英文のコーパスを構築し、文章中の接続詞や句読点の

配置、順序に関する特徴分析に基づいて、英語中級ノンネイティブ（英語を母国語としないが英語能力が中級レベルの人）にとって理解しやすい構造を持った英語文章を生成するシステムを開発し、その有効性を実験的に示したもので、得られた成果は以下の通りである。

- (1) 国際語として英語が広く普及するにつれ、ノンネイティブが理解しやすい文章を生成することが重要となっていることを指摘し、ノンネイティブ向け英文コーパスを構築し、接続詞および句読点の配置を規定する特徴を分析することによって、ノンネイティブ向け文章の持つ修辞構造上の特徴を明らかにした。
- (2) 修辞構造が与えられた場合、接続詞および句読点のそれぞれに対して、文章内での配置の適切性をランク付けするためのルールをコーパス分析に基づいて設計した。
- (3) コーパス分析から得られた接続詞、句読点配置ルールに基づいて文章構造の読み易さを評価する評価関数を組み込んだ遺伝的アルゴリズムによって、修辞構造からノンネイティブ向け英語文章を生成するシステムを開発し、その有効性を実験的に示した。

以上本論文は、数億人にも上るといわれる英語中級ノンネイティブにとって読み易い英語文章を生成することを目的として、ノンネイティブ向け英文コーパスの構築と特徴分析、その結果に基づいた文章生成システムの開発を行い、実験的にその有効性を示したもので、学術上、実際上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。

また、平成20年2月22日実施した論文内容とそれに関連した試問の結果合格と認めた。