氏　　名	：	山崎　誠
専攻分野の名称	：	博士（学術）
学 位 記 番 号	：	博甲第 248 号
学位授与年月日	：	平成２７年３月１７日
学位授与の要件	：	学位規則第４条第１項該当　課程博士
学 位 論 文 名	：	テキストにおける語彙的結束性の計量的研究
論文審査委員	：	（主査）　教授　齋藤　ひろみ
		（副査）　教授　浅沼　茂　　教授　伊坂　淳一
		教授　金澤　裕之　　教授　大井田　義彰

学 位 論 文 要 旨

本研究は、日本語のテキストを対象として語彙的結束性（lexical cohesion）が語彙の量的な側面にどのように現れるかを明らかにすることを目的とする。具体的なリサーチ・クエスチョンは以下の 3 点である。

１．語彙的結束性はテキスト全体の語彙の計量的特性とどのような関係にあるか。

２．語彙的結束性はテキスト中の語彙の分布にどのような形で現れるか。

３．語彙的結束性はテキストの構造にどのように関係しているか。

第 1 章では、語彙的結束性の概念を検討し、一貫性との関係を位置付けた。また、先行研究を紹介するとともに、利用するデータ（『現代日本語書き言葉均衡コーパス』、BCCWJ）とそのアノテーションについて紹介した。

第 2 章ではテキスト全体の計量的特徴が語彙的結束性にどのように係わるかを示した。

第 1 節では、従来語彙の計量的指標としてとらえられてきた異なり語数の延べ語数に対する比、すなわちテキストにおける 1 語当たりの平均使用度数からそのテキストの持つ特徴（特に文章論的な観点から見た特徴）をとらえた。その結果、平均使用度数の高いテキスト及び低いテキストには文体的な特徴があることを指摘した。

第 2 節では、共起語集合（キーとなる語の前あるいは後ろの特定の位置に出現する語の集合）という考えを用いて、BCCWJ においてコロケーションが現れる様子を計量的な指標の観察から記述した。

第 3 節では、語彙の豊かさを表す指標として知られている TTR（Type/Token Ratio）の値がテキスト中で使用されているどの品詞の影響を受けているか計量的に調査した。語数をほぼ一定にしたデータをコーパスから抽出し、当該テキスト全体の TTR と各品詞の TTR との相関を調べたところ、名詞類が一番相関が高く、動詞類、形容詞類がそれに次ぐことが分かった。また、名詞類の中でも普通名詞との相関が高く、TTR の値は普通名詞の使用に大きく左右されることが確認された。

第 3 章ではテキストにおける語の分布から語彙的結束性がどのように働いているか、その一面を明らかにした。

第 1 節では、テキストにおける見出し語の出現間隔の分布とレジスターとの関係を考察した。その結果、高頻度語（主に機能語）の出現間隔は、BCCWJ のレジスターによって違いがあることが分かった。また、テキストにおける出現間隔の平均は、レジスターによる違いは見られないが、出現間隔の総個数がレジスターにより違いがあった。

第 2 節では、テキストにおける多義語の語義は 1 つの語義に偏りやすいが、必ずしも強い制約ではないことをコーパスを利用して明らかにした。また、その偏りは出現間隔が短いほど起こりやすいことを指摘した。これらの現象は語彙的結束性がひとまとまりのテキストに対して働いていることの現れであろうと推測される。

第 3 節では、テキストに出現する多義的な名詞の意味が特定の 1 つの意味で用いられやすいことを観察した。普通名詞を観察した結果、テキスト中では約 7 割～8 割の多義語が特定の意味でのみ使用されていることが分かった。その例外となっているのは、ほとんどが文法的な意味での使用に関わるものであり、文法的な意味は語彙的結束性に関与する度合いが低いことを指摘した。

第 4 章では、語彙的結束性がテキストの構造にどのように関係するのかを明らかにした。

第 1 節では、単純な指標である共起語率を用いて文章の結束性の度合いを観察した。その結果、法律、白書、国会会議録のように結束性の高い文章と新聞、ベストセラー、雑誌のように結束性の低い文章があることが分かった。NDC 別に観察したデータでは、文学の結束性が低いという結果になった。これは文学に会話文が多く、その会話が 1 段落と認定されているというデータの特徴の現れである。

第 2 節では段落間の非対称的類似度を利用して、テキストの結束性のようすを概観した。今回扱ったデータは白書のサンプル 1 つのみであったが、すべての段落間の組み合わせを観察することにより、どの段落とどの段落とが関係が深いのか結束性の一端を伺うことができた。また、隣接した段落以外にも結束性の高い段落があり、それらの関係を利用したテキストの構成の分析への発展の可能性を示唆した。

第 3 節では、語彙的結束性の典型的な現象である、同語の繰り返しに基づく、用語類似語を利用して、話題の展開を測る尺度を提案し、テキスト中の意味段落を切り出す試みを行った。この方法の限界は、算出の方法として意味情報のほとんどを捨象しているため、話題の展開を測る尺度とは言うものの、話題が新しいものに変わったかどうかしか示すことができないことである。文章論で議論される、「順説、逆説、累加、対比」などの関係は分析できない。これは客観的な測定のため、意味情報を積極的に使用しなかったことの代償である。

以上の考察を通して、日本語のテキストにおいて語彙的結束性が定量的に捉えられるいくつかの側面を具体的に明らかにした。また、語彙的結束性は、テキストにおける普遍的な性質として現れるだけでなく、個々のテキストにおいてはそのテキストの個性を表す属性としても現れることも示した。