The University of Osaka
Institutional Knowledge Archive

Title	中国語作文における言語特徴の考察：多次元分析による日本人中国語学習者と中国語母語話者の作文の比較
Author(s)	徐, 勤
Citation	大阪大学, 2023, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/91825
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

様式３

論　文　内　容　の　要　旨

氏　名　（　　徐　勤　　）	
論文題名	中国語作文における言語特徴の考察： 多次元分析による 日本人中国語学習者と中国語母語話者の作文の比較

論文内容の要旨

本論文では，日本人中国語学習者の中国語作文と中国語母語話者の作文とを比較し，多次元分析（Multi-features / Multi-dimensional Analysis）により作文コーパスに見られる言語変異（Linguistic Variation）の相を明らかにし，中国語教育の視点から考察を加えた。まず，因子分析（Factor Analysis）を通じて，学習者によって書かれた作文における言語変異を反映する次元を特定した上で，量的分析と質的分析を相補的に実施し，同一次元における学習者作文と母語話者作文間の比較，及び次元を構成する各言語項目の使用状況を分析することにより，学習者の中国語作文における言語特徴と問題点の特定を可能にした。さらに，学習者が使用した教科書の作文例文における言語表現を調査することで，学習者の産出した書き言葉の特徴及び問題点の要因を明らかにした。また本研究の成果に基づき，中国語作文指導の改善へ向けての提言を行った。これにより，日本人中国語学習者の中国語作文スキル，さらには文章構成力の改善につながることが期待される。

本論文は全9章から構成される。各章の主な内容を以下にまとめる。

第1章「序論」では，本研究の背景と目的，論文の構成，使用するデータとツール及び主な分析手法を紹介する。

第2章「先行研究」では，言語変異の研究分野における多次元分析の先行研究を整理する。まず，多次元分析の定義を説明する。次に，英語の言語変異研究と中国語の言語変異研究における多次元分析の応用を概観する。最後に，本研究の位置付けを示す。

第3章「言語項目の特定」では，説明変数として多次元分析に用いる言語項目（Linguistic Features）を定める。多次元分析の先行研究で言及された言語項目を整理した上で，本研究の中国語言語項目を選定する。本研究では，多次元分析に関する先行研究，中国語の語彙や文法研究に関する専門書，及びNLPIR（中国語の単語分割と品詞解析システム）のタグセットなどを参考にして，名詞・動詞・形容詞・副詞・複文などの10カテゴリーに分類される92種の中国語言語項目（例，抽象名詞，具象名詞，心理動詞，一人称単数代名詞“我”，語彙多様性，平均文長など）を特定した。最後に，テキストにおける中国語言語項目の出現頻度を集計するために開発したコンピュータプログラムを紹介する。本研究で用いた分析プログラムは，オンラインで中国語作文テキストをアップロードすることにより，コーパスを構成する各テキストにおける92種の言語項目の粗頻度（Raw Frequency）と相対頻度（Relative Frequency）を自動的に集計することができ，集計した結果をCSVファイルとして保存することが可能である。

第4章「コーパスの構築」では，本研究で収集した日本人中国語学習者の作文テキストを用いて学習者の作文コーパスを構築する。本章では，コーパスの基本情報を紹介してから，コーパスの検索機能及び検索手順と検索結果の実例を用いて説明する。本研究で構築したコーパスには，キーワードによるセンテンスの検索，品詞・ワードリスト・コンコーダンスの検索などの機能を設けている。キーワードによるセンテンス（Keywords in Sentence）の検索機能は，キーワードが含まれるセンテンスを自動的に抽出することができ，最大5つのキーワードによるセンテンスの検索ができる。品詞（Part-of-Speech）とワードリスト（Wordlist）の検索機能はそれぞれ，ジャンルごとに特定の品詞（例，人称代名詞，副詞）や単語リストに属する語彙（例，低難度語彙 / HSK1-2級の語彙）を抽出し，その品詞や単語リストに対応する語彙の粗頻度と相対頻度（1,000字あたりの出現頻度）を自動的に頻度順に並べ替えた上で，抽出された語彙をWord cloudやBar plotで可視化することもできる。コンコーダンス（Concordance）検索の機能は，指定したキーワードの前後の文脈を抽出することができ，またキーワード前後の文脈の長さをSlider barで指定することが可能である。これらの検索機能により，キーワードを含むセンテンスの抽出，語彙の使用状況，サブコーパス間の比較などが容易に行える。

第5章「多次元分析」では，多次元分析の主な手法である因子分析の手順，データの処理，分析の結果，共起パターンの特定，次元スコアの計算などについて詳述する。本章では，Pythonの「Factor_analyzer」パッケージにおけるミンレス法（Minres Method）という因子の抽出方法，及びプロマックス回転（Promax Rotation）という因子軸の回転方法

を用いて因子分析を行った結果，言語変異を反映しうる4つの因子（言語項目の共起パターン）を特定することができた。本章の最後では，次元スコアの算出方法を紹介するとともに，Pythonプログラミングによって自動算出した各テキストの次元スコアを使用することで，サブコーパス毎の次元スコアの平均値を示した。

第6章「中国語作文における言語変異の次元」では，第5章で得られた4つの因子を主な言語変異の次元として捉え，各次元を構成する有効な言語項目の伝達機能を吟味しながら解釈し，ジャンル間の類似点と相違点を考察する。各次元において高いスコアを示す言語項目が共有する伝達機能に基づき，4つの次元をそれぞれ，「次元1：産出語彙の豊富さvs. 産出語彙の限定性」，「次元2：状態状況表現の適切さ」，「次元3：副詞の修飾」，「次元4：動作の描写」とラベリングした。さらに，各作文ジャンルの次元スコアを用いて多重比較を行った。その結果，母語話者の作文ジャンルにおいて，「意見文」は最も特異性を示すジャンルであることが確認された。すなわちほかのジャンルに比べて，「意見文」は産出語彙が最も豊富である一方，状態状況表現，副詞の修飾，及び動作の描写が著しく少ないことが明らかになった。最後に，同一ジャンルにおける学習者作文と母語話者作文間の類似点と相違点を考察した結果，学習者が作文に用いる語彙項目は主に低難度語彙に限定され，産出語彙の豊富さと状態状況表現の適切さの点で母語話者に比べて有意に下回っており，副詞の修飾も少ない傾向があることが明らかになった。

第7章「次元の言語項目におけるジャンル間の比較」では，各次元を構成する言語項目に基づき，同一ジャンルの学習者作文と母語話者作文における言語特徴の相違について論じる。言語項目の出現頻度（相対頻度）を用いて多重比較を行った結果，学習者の作文には，言語項目の過剰使用と過少使用の問題があることが確認された。

第8章「中国語教育への応用」は，第6章と第7章の考察結果を踏まえ，学習者の作文における言語項目の使用特徴や傾向をまとめ，中国語教育に本研究の成果がいかに適用されうるか提言を行う。学習者の作文における特徴や傾向は3点に集約される。一，学習者の作文で使用された語彙は主に難度の低い語彙に限られており，「一人称単数代名詞“我”」・「低難度語彙」・「ぼかし言葉」・「程度を表す副詞」・「接続詞」の使用頻度については，学習者作文が母語話者作文より著しく高い。二，「高難度語彙」・「非HSK語彙」・「時態助詞“着”」・「時間を表す副詞」の使用頻度，及び作文の「平均文長」と「語彙多様性」に関しては，学習者作文が母語話者作文より有意に低い。三，「人物描写」という作文ジャンルでは，学習者は，「動作動詞」・「方向動詞」・「高頻度動詞」の過少使用が認められる。列挙した特徴と問題点は，主に母語である日本語の干渉，教科書に現れた語彙の影響などに帰されると結論づけた。本章の最後に，中国語教育における作文指導のあり方について，次の3点の提言を行った。一，人称代名詞の表現ストラテジー及びゼロ代名詞の習得に重点を置くこと。二，学習者には，短いセンテンス（短文）を統合し，長いセンテンス（長文）を組み立てることができるようなトレーニングを提供することが必要であること。三，学習者の母語を考慮した上で，語彙（例，程度副詞，接続詞）の導入や教科書に掲載する語彙の選定に大いに工夫の余地があることを主張している。

第9章「結び」では，まず，各章の関連性をまとめた上で，本論の研究意義を述べ，本研究の限界及び今後取り組むべき課題について述べる。本論の不足点及び今後の課題は大別して3点ある。1点目は，次元解釈の問題である。本論文では，量的手法（因子分析）で客観的に抽出された次元は，主に質的分析の視点から解釈してラベリングされたものであるため，ラベリングから全て恣意性を排除することはできない。2点目は，作文コーパスの均衡性の問題である。本論では，サブコーパスのサイズに不均衡がある面は否めない。3点目は，中国語作文教育向けの提案である。本研究では，学習者の中国語作文における特徴や問題点を基に，中国語教育の観点から，どのような内容を教えるべきか，或いは，どのような教育内容に重点を置くべきかについて，具体的な提案を行うことができたが，どのように教えるべきかを論じることは本研究の範囲外である。上記の3点は今後の課題としたい。

本論の研究意義は次の4点である。1）本研究で構築した作文コーパスは，今後の中国語教育研究，特に日本人中国語学習者により産出された書き言葉を研究対象とする，在日中国語教育研究のための作文データとして提供できると考えられる。2）本研究の考察により解明した日本人中国語学習者の作文における特徴や問題点，そしてそれらに基づく中国語教育に関する提案は，今後，日本人学習者向けの中国語教育，特に中国語作文指導の参考になりうると考える。3）量的分析と質的分析とを組み合わせて，中国語母語話者の作文との比較を行い，日本人中国語学習者の作文における言語変異を明らかにすることで，L2中国語作文における言語変異研究に多次元分析が有効であることを示した。4）多次元分析で特定した言語変異を反映する次元に基づき，各テキストに付与された次元スコアを参考に，作文点数との相関を調べ，次元スコアから作文得点を推定する最尤回帰分析を行うことで，コンピュータによる中国語作文自動採点システム開発に一石を投じることになると考えられる。

様式 7

論文審査の結果の要旨及び担当者

氏　名　（徐 勤）			
論文審査担当者		（職）	氏　名
	主 査	教授	田畑 智司
	副 査	准教授	田中 智行
	副 査	教授	瀧田 恵巳

本博士学位請求論文は，多項目・多次元分析法(Multi-feature / Multi-dimensional Analysis，以下 MF/MD 法)により，日本人中国語学習者の中国語作文と中国語母語話者の作文とを比較し，作文コーパスに見られる言語変異（Linguistic Variation）の相を明らかにするとともに, 分析結果を中国語教育の視点から考察した研究である。MF/MD 法は，因子分析(Factor Analysis)を用いて，極めて多数の言語項目間の相互関係，共起関係を特定することにより，分析対象のコーパスに見られる言語変異に実質的に寄与している言語項目をあぶり出し，言語変異を反映するそれぞれの次元ごとにテクスト間の関係性を視覚化して比較する手法である。本論文の主たる学術的貢献は，MF/MD 法という量的分析と，作文テクストの緻密な読みに基づく質的分析を相補的に組み合わせることで，主要な言語変異の次元を特徴づける言語項目の使用実態を明らかにし，学習者の中国語作文における言語特徴および問題点を詳細に論じている点である。併せて，学習者が使用した教科書の作文例文を分析し，教科書と学習者の産出した書き言葉の特徴との関係についても論じており，本論文で提示された知見が，今後の中国語作文教育の重要な基礎資料となりうることも示している。本論文は全 9 章で構成されており，各章の主な内容を以下に示す。

第 1 章は，研究の背景と目的，論文の構成，使用するデータと分析に用いたツール及び解析手法を紹介している。

第 2 章では，MF/MD 法の定義を示した上で，同手法が用いられている先行研究を概観し，特に英語および中国語の言語変異研究において MF/MD 法がもたらした知見や課題をまとめ，本研究の位置付けを明確にしている。

第 3 章では，説明変数として MF/MD 法に用いる言語項目（Linguistic Features）を定めるために，先行研究で言語変異研究に有効であることが示された 92 点の中国語言語項目（例，抽象名詞，具象名詞，心理動詞，一人称単数代名詞“我”，語彙多様性，平均文長など）を提示している。これら 92 項目はさらに，名詞・動詞・形容詞・副詞・複文などの 10 カテゴリーに分類することができる。説明変数として考察する 92 の言語項目のアノテーションはすべてタグの形で実装され，著者自身が開発した Python のプログラムを用いてテクストにおける出現頻度を粗頻度（Raw Frequency）および相対頻度（Relative Frequency）で集計した結果を示している。

第 4 章は，本論文で使用した日本人中国語学習者ならびに中国人母語話者の作文テクストを集積したコーパスと，コーパス分析のために開発したツール（Keywords in Sentence 検索，品詞・単語リスト複合検索，Word cloud や Bar plot による可視化，コンコーダンス検索）によって，どのようなデータが取得可能かを示している。

第 5 章では，MF/MD 法の手順，データの処理，解析結果の表示，共起パターンの特定，次元スコアの計算などについて詳述されている。本論文では, Python の「Factor_analyzer」パッケージ実装の因子の抽出方法（Minres Method），及び因子軸の回転法（Promax Rotation）という条件を定めて因子分析を行った結果，言語変異を反映しうる 4 つの主要因子（言語項目の共起パターン）を特定し，平均値を 0，標準偏差を 1 に変換するよう尺度化した数値を合計し，次元スコア算出を行い，サブコーパス毎の次元スコアの分布を示してデータの全体像を明らかにしている。

第 6 章では，第 5 章で述べた 4 つの因子を主な言語変異の次元として捉え，各次元を構成する有効な言語項目に通底する伝達機能を質的に解釈し，ジャンル間の類似点と相違点を明らかにしている。各次元において高いスコアを示す言語項目が共有する伝達機能に基づき，4 つの次元はそれぞれ，「次元 1：産出語彙の豊富さ vs. 産出語彙の限定性」，「次元 2：状態状況表現の適切さ」，「次元 3：副詞の修飾」，「次元 4：動作の描写」と定義されている。さらに，各作文ジャンルの次元スコアの多重比較を行った結果，母語話者の作文ジャンルの中で「意見文」は産出語彙が最も豊富である一方，状態状況表現，副詞の修飾，及び動作の描写が著しく少ないなど他のジャンルと比べ最も特異性の高いジャンルであることが確認されている。ジャンルごとに学習者作文と母語話者作文間を比較した結果，学習者が用いる語彙項目は主に低難度語彙に限定されていること，産出語彙の豊富さと状態状況表現の

適切さの点で母語話者に比べて有意に下回っていること，副詞の修飾も過少使用の傾向があることなどを指摘している。

第 7 章では，各次元を特徴づける言語項目に基づき，同一ジャンルの学習者作文と母語話者作文間の言語特徴の相違をもとめ，学習者の作文に共通する言語項目の過剰使用と過少使用を明らかにしている。

第 8 章では，学習者の作文における言語項目の使用特徴や傾向をまとめ，中国語教育に本論文の知見がいかに適用されうるか提言を行なっている。学習者の作文における特徴や傾向は次の 3 点に集約される。1, 学習者の作文で使用された語彙は主に難度の低い語彙に限られており，「一人称単数代名詞“我”」・「低難度語彙」・「ぼかし言葉」・「程度を表す副詞」・「接続詞」の使用頻度については，学習者作文が母語話者作文より著しく高いこと。2,「高難度語彙」・「非 HSK 語彙」・「時態助詞“着”」・「時間を表す副詞」の使用頻度，及び作文の「平均文長」と「語彙多様性」に関しては，学習者作文が母語話者作文より有意に低いこと。3,「人物描写」という作文ジャンルでは，学習者は，「動作動詞」・「方向動詞」・「高頻度動詞」の過少使用が認められること。本論文ではこうした差異は，主に母語である日本語の干渉，教科書で使用されている語彙の影響などに帰されると論じた上で，中国語教育における作文指導のあり方について，次の 3 点の提言を行っている。1, 人称代名詞の表現ストラテジー及びゼロ代名詞の習得に重点を置くこと。2, 学習者には，短いセンテンス（短文）を統合し，長いセンテンス（長文）を組み立てることができるようなトレーニングを提供することが必要であること。3, 学習者の母語を考慮した上で，語彙（例，程度副詞，接続詞）の導入や教科書に掲載する語彙の選定に大いに工夫の余地があること。

最後に，第 9 章では，各章の要点とそれぞれの関連性を示し，本論文の意義を主張するとともに，本研究の限界ならびに今後取り組むべき課題を示している。

総括すると，本論文の貢献は次の 4 点であるといえる。

1）構築した作文コーパスおよび分析のための Web アプリケーションは，今後の中国語教育研究，特に非母語話者による中国語書き言葉を対象とする研究にも再利活用が可能な学術資産となること。

2）日本人中国語学習者の作文における特徴や問題点に関する知見，そしてそれに基づく中国語教育に関する提案は，今後，日本人学習者向けの中国語教育，特に中国語作文指導の価値ある基礎資料となること。

3）量的分析，視覚化と質的分析とを組み合わせて，中国語母語話者の作文との比較を行い，日本人中国語学習者の作文における言語変異を明らかにすることで，L2 中国語作文の習熟度研究に MF/MD 法が有効であることを示している点。

4）MF/MD 法で特定した言語変異の次元に基づき，各テクストに付与された次元スコアと作文点数との相関を調べ，次元スコアから作文得点を推定する最尤回帰分析を行うことで，コンピュータによる中国語作文自動採点システム開発に寄与しうること。

本論文が用いた MF/MD 法は，大量のテクストデータから有用な情報を掘り起こすことを支援し，分析者が巨視的視点で発見した現象を微視的に考察する糸口を与えてくれる。この手法を学習者作文の分析に効果的に活かすには，量的分析によりデータの全体像を把握する一方，個々のテクストの特徴を具に捉え，マクロな視点で立てた仮説をミクロな分析や個々のデータの質的解釈により検証し，それをまたマクロな分析にフィードバックし，さらに仮説を精緻化していくという仮説発見と検証の循環を築くことが不可欠である。徐勤氏の論文は，その点において，定量的手法と分析結果の質的解釈をうまく組み合わせて，日本人学習者による中国語作文の特徴を手堅く記述しており，客観性，再現性や反証可能性を担保するための科学的な手順を踏まえた研究となっている。本博士学位請求論文は，日本人中国語学習者による作文研究に新たな知見をもたらした価値ある研究であると評価できる。

一方で，特定した言語変異の次元の解釈は，特徴的言語項目の機能的解釈に依存しており，解釈の恣意性をすべて排除することができているわけではない。また，サブコーパスの構成に若干の不均衡が認められるため，今後新たなデータを取得して補正することが必要であろう。最後に，得られた知見を基に，中国語教育の観点から，教育内容改善のための提言も行われているが，提言の有効性を検証するには，別の研究を待たねばならない。しかし，これらは本論文の学術的価値を損なうほどの問題点ではなく，全体として本論文は，MF/MD 法によるテクスト分析を行なって日本人学習者による中国語作文の特徴を多角的に論じた意義深い研究であり，その学術的貢献は明確である。

以上により，本論文を博士（言語文化学）の学位論文として価値のあるものと認める。