

Title	日本人中国語学習者の叙述文における言語的特徴の分析 : MF/MD 法を使って
Author(s)	徐, 勤
Citation	言語文化共同研究プロジェクト. 2021, 2020, p. 27-41
Version Type	VoR
URL	https://doi.org/10.18910/85007
rights	
Note	

Osaka University Knowledge Archive : OUKA

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

Osaka University

日本人中国語学習者の叙述文における言語的特徴の分析 —MF/MD 法を使って—

徐 勤

大阪大学言語文化研究科
〒560-0043 豊中市待兼山町 1-8
Email: u099322c@ecs.osaka-u.ac.jp

概要 本研究では、中国語母語話者の作文データを比較対象として取り上げ、多項目・多次元法（MF/MD 法）の分析視点から、日本人中国語学習者の叙述文における言語的特徴を考察する。まず、探索的因子分析で言語項目間の共起パターンを抽出し、統計的に有効な因子を7つ採用した。そして、7つの因子を言語変異の主な次元として捉え、それぞれを「書面的 vs. 口語的」、「詳細情報の叙述」、「文脈的複雑性」、「個人の意見」、「過去の叙述」、「ロジックの繋がり」及び「行為の叙述」とラベリングした。作文データの例文を組み合わせて分析した結果、日本人中国語学習者が書いた叙述文には、口語的な傾向があり、文脈的複雑性とロジックの繋がりが不足している一方で、単純な動詞で動作を描写する傾向と、過去の出来事を現在形で叙述してしまう問題点があるということも明らかになった。

キーワード 日本人中国語学習者, MF/MD 法, 言語的特徴

日本人汉语学习者记叙文写作语言特征研究 —基于MF/MD分析法—

徐 勤

大阪大学言語文化研究科
〒560-0043 日本大阪府豊中市待兼山町1-8

摘要 本研究以汉语母语者的作文语料作为比较对象，从多维度·多特征（MF/MD）方法的分析角度出发，考察了日本汉语学习者记叙文写作的语言特征。首先，通过探索性因子分析提取出语言项目之间的共现模式，得到了7个具有统计学意义的有效因子。其后，将7个因子作为语言变异的主要维度，分别标注为“书面性与口语化”、“叙述细节信息”、“语句复杂性”、“表述个人意见”、“叙述过去事件”、“逻辑关联性”、“描述动作行为”。最后，结合作文文本的例句进行定性分析，结果显示，日本汉语学习者的记叙文具有口语化倾向，语言表达缺乏复杂性和逻辑关联性，同时，学习者还偏好于使用简单的动词进行动作描写，以及误用一般现在时态来描述过去事件的问题。

关键词 日本人汉语学习者, MF/MD分析法, 语言特征

1. はじめに

MF/MD 法 (Multi-Features / Multi-Dimensional Approach : 多項目・多次元法) とは、主に因子分析を使用することで、言語項目間 (変数間) の複雑な共起関係 (co-occurrence) を明らかにし、言語変異を多次的に捉える手法である。この分野では、Douglas Biber の研究成果は注目に値する。Biber (1988) は書き言葉の LOB コーパス (Lancaster/Oslo-Bergen Corpus) と話し言葉の LLC コーパス (London-Lund Corpus) を基に、新聞報道・論説・伝記・政府刊行物・学術散文・推理小説・ユーモア・即席スピーチ・準備したスピーチ等の 23 ジャンル (合計 481 個のサンプル) のテキストにおける 67 種の言語項目の出現頻度を集計し、因子分析によって言語変異を解釈できる 7 つの因子を次元 (Dimensions) として捉えている。

次元 1: Involved versus Informational Production

次元 2: Narrative versus Non-narrative Concerns

次元 3: Explicit versus Situation-Dependent Reference

次元 4: Overt Expression of Persuasion

次元 5: Abstract versus Non-Abstract Information

次元 6: On-Line Informational Elaboration

次元 7: Academic Hedging

しかし、次元 7 は言語項目数も統計的な有意性も非常に小さいため、実際の分析には使用されていない。Biber (1988) では「最終的に 6 因子が次元として提案されている」(McEney & Hardie 2011:106 (石川 2014:159) 訳)。さらに、大学の言語使用域 (university registers) における英語の口語 (例: Office hours, Study groups 等) と書面語 (例: Textbooks, Institutional writing 等) の区別を明確にできるように、Biber (2006) は T2K-SW AL コーパス (the TOEFL 2000 Spoken and Written Academic Language Corpus) を用い、MF/MD 法によって抽出した 4 つの因子を言語変異の 4 つの次元と解釈している。

次元 1: Oral vs. literate discourse

次元 2: Procedural vs. content-focused discourse

次元 3: Reconstructed account of events

次元 4: Teacher-centered stance

Biber の研究成果は言語変異の研究分野に大きく貢献している。田畑 (2005) は「Biber の言語変異研究モデルは使用域の問題だけでなく散文文体の通時的変化」、「英語以外の言語における言語変異の研究」(p.192-193) にも適用されると述べている。また、Biber の研究モデルは外国語学習にも応用されている。例えば、Friginal & Weigle (2014) は学習者の英作文の特徴を解釈するために、Biber (1988) の視点に基づき、英語学習者の作文を分析データとして取り上げ、探索的因子分析 (Exploratory Factor Analysis) によって 4 つの次元を抽出し、L2 ライティングのマイクロ変化を分析している。

次元 1: Involved vs. Informational Focus

次元 2: Addressee-Focused Description vs. Personal Narrative

次元 3: Simplified vs. Elaborated Description

次元 4: Personal Opinion vs. Impersonal Evaluation/Assessment

近年では、MF/MD 法を使って、中国語の書き言葉と話し言葉の言語使用域に関する言語変異の研究 (Zhang, 2012; Zhu, 2015; 劉, 2019) が見られるが、管見の限りでは、中国語教育分野では、MF/MD 法による中国語学習者の作文における言語的特徴に関する研究が見当たらない。そのため、これを研究テーマとして取り組みたい。

2. 方法

2.1. 目的

中国語ネイティブスピーカーの書き言葉と比べて、日本人中国語学習者が書いた中国語作文にはどのような言語的変異が見られるのかを、MF/MD 法で考察する。

2.2. 研究対象とコーパス

本研究では、日本人中国語学習者が書いた叙述文を研究対象として取り上げる。具体的には、大阪大学の中国語専攻に在籍しており、上級中国語作文授業を受講している大学3年生の日本人が書いた中国語作文を収集して電子化し、学習者（以下は「JP」とする）の作文コーパスにしたものを分析データとする。JP コーパスには、260 篇の叙述文テキストが含まれている。さらに、学習者の作文における言語的特徴と比較対照するために、本研究では中国国内の高校3年生¹が書いた叙述文（260 篇）も収集し、母語話者（以下は「CHN」とする）の作文コーパスにしたものを比較対象とする。

表 1 コーパスの基本情報

	N (総篇数)	総字数	総語数	異なり語数
JP	260	145,959	100,244	45,221
CHN	260	249,599	182,053	82,543

2.3. 言語項目

変数として用いられる言語的特徴 (linguistic features) は、名詞・動詞・形容詞・複文などの 21 カテゴリーからなる中国語の 111 種の言語項目 (Appendix I) である。111 種の言語項目における 58 種は Biber (2006) によるもので、主に英語と中国語に共通する言語的特徴 (例えば、心理的動詞/mental verbs、副詞/adverbs 等) である。また、Biber (2006) の 4 つの次元には含まれていない 3 つの言語的特徴 (最もよく使われる名詞/noun: most frequently used; 最もよく使われる動詞/verb: most frequently used; 最もよく使われる形容詞/adjective: most frequently used) も本研究の言語項目

¹ 張(2018)では、“高中三年级学生的汉语水平能够代表汉语母语者的语言水平” (訳: 高校3年生の中国語レベルは、中国語ネイティブスピーカーの言語レベルを代表することができる) と述べているため、本研究は高校3年生の中国人が書いた作文を参照対象として取り上げる。

リストに設けている。中国語には節（例、that-clauses, to-clauses, relative clauses 等）がないものの、複文が用いられているため、黄と廖(2017) で考察されている 10 種類の複文は分析に含まれている。そのほかに、中国語の言語的特徴及びアカデミックと非アカデミック的な中国語に関する研究（Zhu, 2015; Zhang, 2012; 劉, 2019; 馮ら, 2008）を参考にして、40 種の言語項目（例：着/zhe, 了/le, 过/guo, 的/de, 地/de, 得/de, 口語表現(口語詞/oral words), 二音節テンプレートで用いられる単音語(嵌偶単音詞/monosyllabic words used in disyllabic templates), 古典的な中国語単語(古語詞/particular classical Chinese words 等) が言語項目の構成要素として追加されている。

2.4. 分析手順

(1) 単語分割とタグ付け

各コーパス内の作文テキストに対し、NLPIR で単語分割とタグ付けを行う。

(2) 言語項目の頻度集計

各テキストにおける各言語項目の出現頻度を集計し、テキスト毎の頻度表を作成する。そして、出現頻度を 1,000 語当たりの相対頻度に調整する。

(3) 因子分析

相対頻度のデータを使って因子分析を行う。本研究の因子抽出法は Unweighted Least Squares (重みなし最小 2 乗法) であり、因子軸の回転法は斜交回転法における Promax Rotation (プロマックス回転) である。

(4) 次元の解釈

因子分析によって統計的に抽出された各因子を主な次元として捉え、機能ラベルを付与して解釈する。

3. 結果と考察

本研究は111種の言語項目を設けたが、Biber (2006:182-183)² が述べている三つの理由に基づいて、一部分の言語項目を削除し、変数として因子分析に使用した言語項目は58種である。また、因子分析を行う条件として、サンプル数が観察変数の5倍以上になる必要がある (Gorsuch 1983:322; Biber 1988:65)。本研究のテキストの数 (520) と観察変数 (58) は因子分析の条件を満たしている。

因子分析を行う前に、KMOとBartlettの球面性検定でサンプルサイズの妥当性を検証する必要がある。本研究で得られたKMO値は0.756 (Kaiser(1974)の基準³を参照) で、Bartlettの球面性検定

² Biber(2006:182-183):

(1) Some features were dropped because they overlapped to a large extent with other features.

(2) Some features were dropped because they were extremely rare.

(3) Some features were dropped because they shared little variance with the overall factorial structure (features with communalities below 0.15 do not have meaningful factor loadings on any factor).

³ Kaiser(1974):

0.90 = marvelous ; 0.80 = meritorious ; 0.70 = middling ; 0.60 = mediocre ; 0.50 = miserable ; below 0.50 = unacceptable

の有意確立は $p < 0.001$ であったため、因子分析を行う妥当性があることが確認された。

表 2 KMO と Bartlett の球面性検定

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	.756	
Bartlett's Test of Sphericity	Approx. Chi-Square	15658.272
	df	1653
	Sig.	.000

そして、本研究は「説明された分散の合計 (Total Variance Explained)」とスクリー基準(Scree criterion)で7因子解 (図 1) を採用した。最初の7つの因子の累積寄与率 (表 3: 第1因子から第7因子まで順番に加算した寄与率) は約 40.69%であるため、7つの因子でデータ全体の約 40.69%を説明することができる。

表 3 説明された分散の合計

Total Variance Explained							
Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	8.432	14.538	14.538	8.027	13.840	13.840	6.836
2	4.833	8.333	22.870	4.556	7.855	21.694	3.569
3	3.696	6.372	29.243	3.284	5.662	27.356	4.537
4	2.812	4.847	34.090	2.366	4.080	31.436	3.402
5	2.304	3.972	38.062	1.960	3.378	34.814	3.054
6	2.091	3.605	41.666	1.722	2.969	37.783	3.793
7	2.027	3.496	45.162	1.687	2.908	40.691	2.458
8	1.868	3.221	48.383	1.422	2.452	43.144	2.084

表 4 因子間相関行列

Factor	1	2	3	4	5	6	7
1	1						
2	-.26	1					
3	.33	-.38	1				
4	.02	.15	.30	1			
5	.20	-.14	.05	-.07	1		
6	.30	-.03	.48	.28	.19	1	
7	-.13	-.07	.00	.05	.30	.18	1

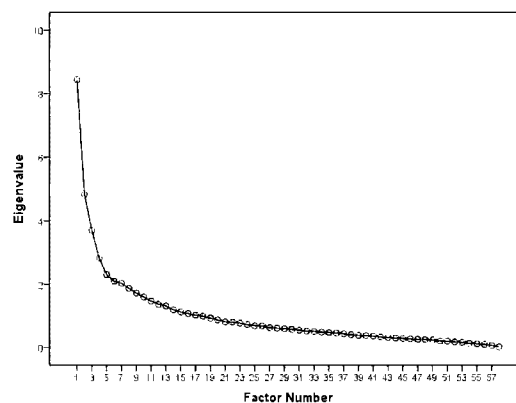


図 1 因子のスクリープロット

表4は因子軸をプロマックス回転した後の因子間相関の結果である。表4の相関係数を見ると、7つの因子は全般的に相互相関が低いことがわかった。つまり、7つの因子それぞれが、言語の変化の独立した次元を反映している。

3.1. 共起パターン

表5は58個の言語項目間の共起パターンである。このパターンは因子軸をPromax Rotationで回転した後のPattern Matrixの結果である。Biber(1988:87)によると、絶対値が0.30未満の負荷量(Loadings)は、たとえ統計的に有意であっても、一般的には重要でないとして除外される。そのため、表5に因子負荷量の絶対値は0.30以上の言語項目しか表示されていない。また、負荷量の右側に括弧を付けている言語項目は、「他の因子にも重複して現れる項目で、他の因子における負荷量の方が高いため、因子得点の計算には使用されない」(田畑, 2002:83)。

表5 言語項目の共起パターン⁴

Factor 1		Factor 3	
词汇多样性 (Lexical diversity)	.77	副词: 最常用 (Adverbs: most commonly used)	.75
助词“着” (Auxiliary: zhe/着)	.67	并列复句 (Parallel compound sentences)	.71
进行式动词 (Progressive verbs)	.65	副词: 表态度 (Adverbs: attitude)	.65
副词: 中度常用 (Adverbs: moderately commonly used)	.61	时间副词 (Adverbs of time)	.47
平均句长 (Average sentence length)	.61	假设复句 (Hypothetical compound sentences)	.41
形容词: 非常用 (Adjectives: rarely used)	.58		
状态词 (State words)	.53	Factor 4	
大小形容词 (Adjectives of size)	.42	副词: 表必然性 (Adverbs: inevitability)	1.02
助词“的” (Auxiliary: de/的)	.42	否定词 (Negative words)	1.00
嵌偶单音词 (monosyllabic words used in disyllabic templates)	.39	Factor 5	
方位词 (Position words)	.36	过去式动词 (Past verbs (e.g., heard that/曾听))	1.01
时间副词 (Adverbs of time)	.33	助词“了” (Auxiliary: le/了)	.90
.....		Factor 6	
因果复句 (Causal compound sentences)	-.47	条件复句 (Conditional compound sentences)	.90
句首介词 (Prepositions in the beginning of a sentence)	-.52	顺承复句 (Successive compound sentences)	.67
第一人代词“我” (First person pronoun: I/我)	-.64	假设复句 (Hypothetical compound sentences)	.47
Factor 2		Factor 7	
合偶双音词 (Couple disyllabic words)	.74	动作行为动词 (Action verbs)	.86
名动词 (Nominal verbs)	.67	动词: 最常用 (Verbs: most commonly used)	.71
平均词长 (Average word length)	.53	趋向动词 (Directional verbs)	.48
集体名词 (Collective nouns)	.39	
动词: 最常用 (Verbs: most commonly used)	.36	助词“的” (Auxiliary: de/的)	-.35
.....			
处所词 (Location nouns)	-.35		

⁴ 因子軸を Promax Rotation で回転した後の Pattern Matrix の結果における因子負荷量は回帰係数 (regression coefficients) であるため、因子負荷量の大きさが1を超えることが可能である (In Factor Analysis (or in PCA), what does it mean a factor loading greater than 1? - Cross Validated (stackexchange.com))。

3.2. 次元の解釈

表5における各因子を構成する言語項目に基づき、各因子の機能を解釈する。ここでは、第1因子から第7因子までを主な次元（Dimensions）として捉える。

3.2.1 次元1：書面的 vs.口語的

次元1のスコア及び次元1を構成している言語項目の相対頻度に、両コーパスの間で有意差があるかどうかについて、本研究は対応のないサンプルのT検定を行った。

表 6 対応のないサンプルの T 検定（次元1）

	JP		CHN		p-value	Effect size
	N	Mean	N	Mean		
Factor score 1	260	-0.83	260	0.83	.000	3.32
词汇多样性 (Lexical diversity)	260	8.81	260	11.98	.000	2.09
助词“着” (Auxiliary: zhe/着)	260	1.13	260	9.42	.000	1.82
进行式动词 (Progressive verbs)	260	4.44	260	12.81	.000	1.45
副词：中度常用 (Adverbs: moderately commonly used)	260	3.52	260	8.23	.000	1.22
平均句长 (Average sentence length)	260	15.29	260	27.00	.000	1.45
形容词：非常用 (Adjectives: rarely used)	260	0.61	260	2.63	.000	0.95
状态词 (State words)	260	0.52	260	4.96	.000	1.43
大小形容词 (Adjectives of size)	260	3.13	260	5.66	.000	0.61
助词“的” (Auxiliary: de/的)	260	55.40	260	60.33	.001	0.30
嵌偶单音词 (monosyllabic words used in disyllabic templates)	260	18.24	260	23.01	.000	0.52
方位词 (Position words)	260	13.57	260	18.27	.000	0.64
时间副词 (Adverbs of time)	260	17.73	260	31.31	.000	1.45
因果复句 (Causal compound sentences)	260	8.81	260	3.29	.000	1.05
句首介词 (Prepositions in the beginning of a sentence)	260	6.57	260	2.19	.000	1.16
第一人代词“我” (First person pronoun: I/我)	260	54.44	260	27.66	.000	1.19

結果としては、すべて有意差 ($p < 0.05$) があるということが明らかになった。また、言語項目の相対頻度の平均値について、正の負荷量を持っている言語項目は、JP（学習者）の方が過少使用し、負の負荷量を持っている言語項目は、JP（学習者）が過剰使用している。次元1では、正の負荷量が一番高いのは語彙多様性（词汇多样性/Lexical diversity）であり、負の負荷量が一番高いのは一人称代名詞“我”（First person pronoun: I/我）である。これは書面的と口語的な対比と言えるので、次元1に「書面的vs口語的」という名前を付ける。

- (1) 我喜欢吃中国菜，特别是回锅肉。我妈妈教我菜谱。我妈妈很会做菜，所以我想像我妈妈。我妈妈的得意菜是干烧虾仁，所以我想一起做。最近我也喜欢做点心。但是，因为失败了，所以我想再做。（JP_051）

文字通りの訳：私は中華料理、特にホイコーローが大好きです。私の母が私にレシピを教えてください。 私の母は料理が得意なので、だから私は私の母のような人間になりたいです。私の母の得意な料理はエビチリで、だから私は一緒に作りたいです。最近

私は点心を作るのも好きです。しかし、失敗したので、だから私はまた作りたと思います。))

学習者の作文には、一人称代名詞“我”と因果関係（因果復句/Causal compound sentences）を表している接続詞“因为”（～ので），“所以”（だから～）等が多く使用されている。また、JPコーパスにおける語彙多様性（词汇多样性/Lexical diversity）、平均文長（平均句長/Average sentence length）等は、CHN（母語話者）コーパスより低いので、学習者の作文は口語的な傾向があるということが明らかになった。

3.2.2 次元2：詳細情報の叙述

次元2では、因子得点について、両コーパスの間に有意差があり、学習者コーパスの方が因子得点が高い。次元2に属している言語項目の相対頻度を見ると、正の負荷量を持っている言語項目は学習者の方が過剰使用している。負の負荷量を持っている言語項目は、両コーパスの間には有意差がないため、次元2を解釈する際、主に正の負荷量を持っている言語項目で解釈する。

表 7 対応のないサンプルの T 検定（次元 2）

	JP		CHN		p-value	Effect size
	N	Mean	N	Mean		
Factor score 2	260	0.39	260	-0.39	.000	1.05
合偶双音词 (Couple disyllabic words)	260	17.19	260	12.47	.000	1.16
名动词 (Nominal verbs)	260	16.23	260	9.45	.000	1.19
平均词长 (Average word length)	260	1.45	260	1.37	.000	1.05
集体名词 (Collective nouns)	260	11.71	260	5.44	.000	1.16
动词：最常用 (Verbs: most commonly used)	260	142.84	260	127.92	.000	1.19
处所词 (Location nouns)	260	6.84	260	7.27	.392	0.08

しかし、最もよく使われる動詞（动词：最常用/verb: most frequently used）という言語項目は他の因子にも重複して現れる項目（表5）で、次元2の解釈に使用することができない。そのため、最初の4つの言語項目で次元2を解釈してみると、学習者の作文には、学習者の日常生活に関係がある集合名詞（集体名词/Collective nouns）、例えば、「大阪大学」、及び名動詞（名动词/Nominal verbs）、例えば、「影響」、「留学」などがよく出現しているので、次元2を「詳細情報の叙述」と名付ける。

(2) 我有一个中国朋友，是大连理工大学的学生。她去年暑假在大阪大学学了四周日语。那时我当她的志愿者，帮她学习日语。她打算从十月开始在大阪大学一年的留学。今天她给我发短信。我从她发来的短信中知道了因为受新冠病毒的影响，她的留学期间缩短了。

(JP_068)

文字通りの訳：私は大连理工大学の学生である中国人の友人がいます。彼女は今年の夏、大阪大学で3週間日本語を勉強しました。その時、私は彼女のボランティアとして、日本語の勉強を手伝っていました。彼女は10月から大阪大学で1年間の留学をする予定があります。今日、彼女からメールが来ました。彼女のメールによると、新型コロナの影響で留学期間が短縮されたそうです。

3.2.3 次元3：文脈的複雑性

次元3では、因子得点について、両コーパスの間に有意差があり、学習者コーパスの方が因子得点が高い。最もよく使われる副詞（副詞：最常用/Adverbs: most commonly used）のほかの言語項目、例えば、接続詞“一边…一边…”（～しながら～する）を表している並列複合文（並列複句/Parallel compound sentences），“马上”（すぐ）を表している時の副詞（時間副詞/Adverbs of time）などは、学習者の方が過少使用している。母語話者による例文を基に、次元3を「文脈的複雑性」と名付ける。つまり、学習者の作文には、文脈的複雑性が足りないと言える。

表 8 対応のないサンプルの T 検定（次元 3）

	JP		CHN		p-value	Effect size
	N	Mean	N	Mean		
Factor score 3	260	-0.43	260	0.43	.000	1.04
副詞：最常用 (Adverbs: most commonly used)	260	63.34	260	63.82	.751	0.03
並列複句 (Parallel compound sentences)	260	2.89	260	5.57	.000	0.73
副詞：表態度 (Adverbs: attitude)	260	8.71	260	12.60	.000	0.59
時間副詞 (Adverbs of time)	260	17.73	260	31.31	.000	1.45
假设複句 (Hypothetical compound sentences)	260	2.94	260	7.29	.000	1.12

(3) 就在这时，汽车靠站了，上来一位抱着婴儿的阿姨，怀里的婴儿还在不停地哭闹，只见她一边哄小孩说“宝宝别哭了，马上就要到家了”，一边掏钱买票。（CHN_001）

訳：その時、バスは駅に引っ張られ、まだ泣いている赤ちゃんを抱えたおばさんがやってきて、「泣かないで、もうすぐ家に着くから」と赤ちゃんをなだめながら、切符を買うためのお金を取り出した。

3.2.4 次元4：個人の意見

次元4では、両コーパス間の因子得点には有意差がない。次元4に属している言語項目を吟味して、次元4を「個人的意見」と名付ける。

表 9 対応のないサンプルの T 検定（次元 4）

	JP		CHN		p-value	Effect size
	N	Mean	N	Mean		
Factor score 4	260	-0.02	260	0.02	.582	0.05
副詞：表必然性 (Adverbs: inevitability)	260	11.64	260	13.34	.008	0.24
否定詞 (Negative words)	260	10.29	260	12.23	.001	0.30

3.2.5 次元5：過去の叙述

次元5に属している言語項目は過去を記述する特徴であるため、次元5を「過去の叙述」と名付

ける。

表 10 対応のないサンプルの T 検定 (次元 5)

	JP		CHN		p-value	Effect size
	N	Mean	N	Mean		
Factor score 5	260	-0.21	260	0.21	.000	0.42
過去式動詞 (Past verbs (e.g., heard that/曾听))	260	17.59	260	21.36	.000	0.36
助詞“了” (Auxiliary: le/了)	260	11.49	260	14.59	.000	0.37

(4) 这次我们去鸡肉火锅店。其实，昨天晚上吃了鸡肉鸡蛋盖饭，我**吃腻**鸡肉。(JP_001)

文字通りの訳：今回私たちは鶏鍋のお店に**行きます**。実は昨夜、親子丼を食べました、私は鶏肉に**飽きます**。

(5) 时间一天天**过去**了，如今我**已经**长大，父母**已**步入中年，周围的一切都**发生**了很大的变化，但是唯一没变的是他们仍然吵架。(CHN_258)

訳：日が経って、私は**すでに**大人になり、両親も**すでに**中年になりました。周りのことはだいぶ**変わ**ってきたが、唯一**変わ**っていないのは、両親が相変わらず喧嘩ばかりしていることです。

学習者は、過去の出来事を記述する際、“動詞+了”(～した)を使うべき文脈において、“了”を忘れてしまったケースが多い。例(4)の文は、過去の出来事について述べているため、“这次我们去了鸡肉火锅店(今回私たちは鶏鍋のお店に行きました)”、“我**吃腻**了鸡肉(私は鶏肉に**飽**てしまいました)”とするのが意味的に正しい文である。しかし、動詞の後ろに“了”をつけ忘れることにより、過去の出来事を現在形で叙述してしまっている。また、JPコーパスでは、過去を表す動詞(過去式動詞/Past verbs)、すなわち「時の副詞+動詞」の形(例えば“已经+動詞”/すでに～)で完了した動作を描写するケースが少ない。次元5の因子得点についてのT検定の結果を合わせてみると、学習者の作文には過去の叙述は過少である。

3.2.6 次元6：ロジックの繋がり

次元6に属している言語項目である「条件複文」、「承前複文」、「仮定複文」は全て文脈のロジックを表す言語項目である。次元6を「ロジックの繋がり」と名付ける。

表 11 対応のないサンプルの T 検定 (次元 6)

	JP		CHN		p-value	Effect size
	N	Mean	N	Mean		
Factor score 6	260	-0.41	260	0.41	.000	0.97
条件複句 (Conditional compound sentences)	260	1.00	260	3.13	.000	0.91
順承複句 (Successive compound sentences)	260	1.69	260	5.56	.000	1.20
假设複句 (Hypothetical compound sentences)	260	2.94	260	7.29	.000	1.12

JPコーパスでは、次元6を構成している三つの複文を過少使用しており、学習者の作文はロジックの繋がりが不足しているということが分かった。

(5) 他说：“这虽然只是小事，但是如果没有人站出来指出他的错，那么他就会认为自己没有错，以后他就会一直这样下去的。”对啊！就是因为没有人肯出来指出错误，社会上才依然存在各种不良现象。(CHN_006)

訳：「些細なことですが、誰も彼が間違っていることを指摘しなければ、彼は自分が間違っていないと思い、これからもそうあり続けるだろう」と彼は言いました。その通り。誰も前に出て誤りを指摘しようとしなからこそ、様々な悪事が社会に残ってしまうのです。

3.2.7 次元7：行為の叙述

次元7では、因子得点について、両コーパスの間に有意差があり、学習者コーパスの方が因子得点が高い。学習者の作文には、動作動詞（動作行为动词/Action verbs）、例えば“吃”（食べる）/“出门”（家を出る）、及び最もよく使われる動詞（動詞：最常用/verb: most frequently used）、例えば“去”（行く）/“有”（ある）などのような難度の低い動詞がたくさん使われている。そのため、次元7を「行為の叙述」と名付ける。

表 12 対応のないサンプルの T 検定（次元7）

	JP		CHN		p-value	Effect size
	N	Mean	N	Mean		
Factor score 7	260	0.11	260	-0.11	.010	0.23
動作行为动词 (Action verbs)	260	81.38	260	77.82	.042	0.18
動詞：最常用 (Verbs: most commonly used)	260	142.84	260	127.92	.000	1.19
趋向动詞 (Directional verbs)	260	14.42	260	14.45	.966	0.00
助詞“的” (Auxiliary: de/的)	260	55.40	260	60.33	.001	0.30

(6) 早上我出门的时候，没下雨。不过打完工的时候，已经开下起了毛毛细雨。我没有雨伞，被雨淋湿去石桥，因为今天我们的社团开忘年会。进了十二月我有很多酒会，前天也有打工的酒会。这次我们去鸡肉火锅店。其实，昨天晚上吃了鸡肉鸡蛋盖饭，我吃腻鸡肉。(JP_001)

文字通りの訳：朝、出かけたとき、雨は降っていませんでした。しかし、アルバイトを終えた頃には、すでに小雨が降り出していました。今日は私たちのクラブの忘年会があったので、私は傘を持たずに雨に濡れながら、石橋に行ってきた。12月に入ると飲み会が多くなり、おとといもアルバイトの飲み会がありました。今回私たちは、鶏鍋のお店に行きます。実は昨夜、親子丼を食べました、私は鶏肉に飽きます。

3.3. 可視化

各次元のスコアに基づき、Rader chartsで両コーパスの区別を可視化する。CHNによって書かれた叙述文と比較して、JPの叙述文には、口語的な傾向があり、文脈的複雑性とロジックの繋がりが不足している。また、学習者は、過去の出来事を記述する際、動詞の後に“了”を使うべきところで、“了”を使っていない場合がある。さらに、叙述文で動作を記述する際、学習者は単純な動詞を使用する傾向があることが明らかになった。

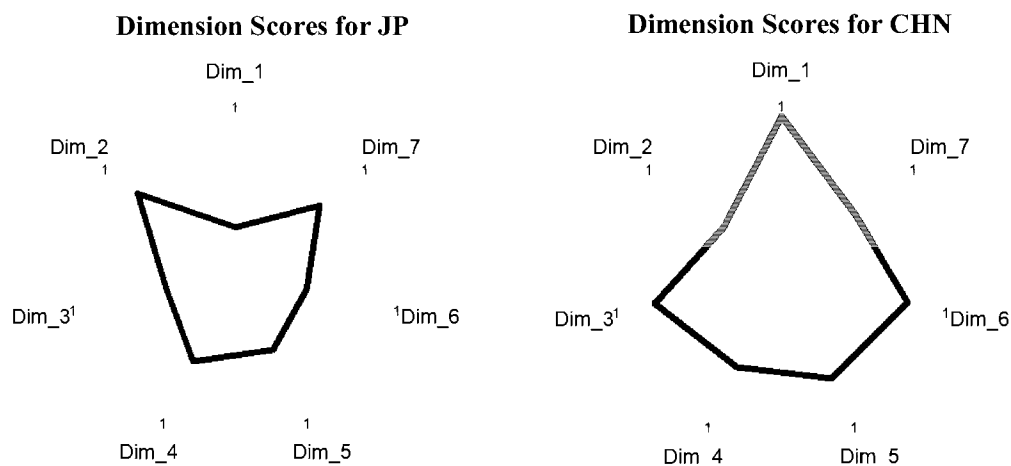


図 2 可視化図

4. 終わりに

本研究は予備的な研究として、MF/MD法の分析視点から、因子分析で7つの次元を抽出し、日本人中国語学習者が書いた叙述文における言語的特徴の考察を行った。

この研究では2つの点が不足していると考えられる。まず、Yong & Pearce(2013) がこの分析方法の制約の一つとして述べているように、因子のラベリングに問題がある可能性がある。つまり、次元を解釈する時に、因子内の変数を正確に反映していない場合がある。また、言語項目の頻度集計に問題があるおそれがある。例えば、中国語の複文はとても複雑であり、本研究では、主に接続詞のタグに基づいて複文の頻度を集計したが、すべての複文を集計できていない可能性もある。

今後の課題としては、上述の2つの不足を補うために、研究手法と分析視点を改善していく。また、コーパスデータを拡大し、ほかの種類 of 作文（例えば意見文、説明文）も収集して分析してみることが必要である。量的分析で得られた結果に基づき、作文テキストにおける典型的な言語表現を抽出し、質的分析を行いながら、日本人中国語学習者の作文における言語的特徴や傾向を明らかにしていきたい。

文献

- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Biber, D. (2006). *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Biber, D. (2014). Using multi-dimensional analysis to explore cross-linguistic universals of register variation. *Languages in Contrast*, 14, 7-34.
- McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Friginal, E., & Weigle, S. (2014). Exploring multiple profiles of L2 writing using multi-dimensional analysis. *Journal of Second Language Writing*, 26, 80-95.
- Gorsuch, Richard. (1983). *Factor Analysis, 2nd edition*. Hillsdale, NJ: Lawrence Erlbaum Association.
- Hardy, J. A., & Friginal, E. (2016). Genre variation in student writing: A multi-dimensional analysis. *Journal of English for Academic Purposes*, 22, 119-131.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36.
- Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology*, 9(2), 79-94.
- Zhu, X. (2015). *A multi-dimensional approach to register variation in Mandarin Chinese* (Unpublished masters' thesis). Zhejiang University, Hangzhou, China.
- Zhang, Z. S. (2012). A corpus study of variation in written Chinese. *Corpus Linguistics and Linguistic Theory*, 8(1), 209-240.
- 田畑 智司. (2002). 「Corpus-Based Stylistics--MF/MD 法による文体比較--」 『英語文体論の方法と射程』 言語文化共同プロジェクト. pp.75-89.
- 田畑 智司. (2005). 「コーパスに基づく文体論研究」 斎藤他 編『英語コーパス言語学：基礎と実践[改訂新版]』 第9章(pp. 183-206) 東京：研究者出版
- 石川慎一郎. (2014). 「(訳) 概説コーパス言語学—手法・理論・実践」 東京：ひつじ書房
- 冯胜利, 王洁, 黄梅. (2008). 汉语书面语体庄雅度的自动测量[An automatic feature checking algorithm for degree of formalities in written Chinese]. *语言科学[Yuyan Kexue]*, 7(2): 113-126.
- 黄伯荣, 廖序东. (2017). *现代汉语 (修订第六版) [Modern Chinese (6th Rev. ed.)]*. 北京: 高等教育出版社 [Beijing: Gaodeng Jiaoyu Chubanshe].
- 刘艳春. (2019). 汉语语体变异的多维度分析——基于 17 个语体 72 项语言特征的考察. [A multi-dimensional analysis to variations in Chinese registers: Based on a study on 72 linguistic features in 17 registers]. *江汉学术[Jianghan Xueshu]*, 38 (3), 100-110.
- 张江丽. (2018). 汉语作为第二语言学习者笔语产出性词汇研究.[A Study on Written Productive Vocabulary of Chinese Second Language Learner]. *世界汉语教学[Shijie Hanyu Jiaoxue]*, 32(3), 417-430.

Appendix I

类别(Category)	语言特征(Linguistic feature)	例(Example)
1 名词及其特殊属类 Nouns & Special Noun Genus	1. 名词: 最常用 (Nouns: most commonly used)	人、社会、问题
	2. 名词: 中度常用 (Nouns: moderately commonly used)	微生物、军官、地主阶级
	3. 名词: 非常用 (Nouns: rarely used)	决赛、西红柿、写字台
	4. 抽象名词 (Abstract nouns)	奥秘、魅力、权术
	5. 立场名词 (Stance nouns)	倾向、意图、信心
	6. 心理名词 (Mental nouns)	悲欢、廉耻、嫌隙
	7. 大学学科专业分类词 (University subject classifications)	哲学、心理学、经济学
	8. 指人名词 (Personal nouns)	书法家、工程师
	9. 集体名词 (Collective nouns)	公社、陆军、编委会
	10. 具象名词 (Figurative nouns)	日记、酸奶、取景框
	11. 具象科技名词 (Figurative scientific and technological nouns)	暗礁、磁场、吞噬细胞
	12. 度量衡名词 (Metrics and measures)	尺寸、价码、储量
2 动词及其特殊属类 Verbs & Special Verb Genus	13. 动词: 最常用 (Verbs: most commonly used)	有、说、到
	14. 动词: 中度常用 (Verbs: moderately commonly used)	送来、被迫
	15. 动词: 非常用 (Verbs: rarely used)	掘、插嘴、试点
	16. 动作行为动词 (Action verbs)	唱歌、操作、挥洒
	17. 使役动词 (Causative verbs)	动摇、落实、复兴
	18. 存现动词 (Existing verbs)	升起、凝聚、失踪
	19. 心理动词 (Mental verbs)	按捺、忏悔、揣摩
	20. 肯定动词 (Affirmative verbs)	认可、查明、领会
	21. 交流动词 (Communicative verbs)	安慰、邀请、叙旧
	22. 推测性动词 (Speculative verbs)	臆想、推测、思索
	23. 可能性动词 (Possibility verbs)	似乎、觉得、看样子
	24. 趋向动词 (Directional verbs)	起、出来、下来
	25. 副动词 (Adverbial verbs)	带头、合力
	26. 形式动词 (Light verbs)	进行、给予、加以
3 形容词及其特殊属类 Adjectives & Special Adj.Genus	27. 形容词: 最常用 (Adjectives: most commonly used)	大、小、多
	28. 形容词: 中度常用 (Adjectives: moderately commonly used)	舒服、最低、得意
	29. 形容词: 非常用 (Adjectives: rarely used)	最优、无用、疲乏
	30. 大小形容词 (Adjectives of size)	大、微小、修长
	31. 常用形容词: 相关性 (Common adjectives: relevance)	大体的、全部的
	32. 状态词 (State words)	悠远、寥寥、活生生
	33. 区别词 (Distinguishing words)	特定、负面、崭新
4 数词和量词 Numerals & Quantifier	34. 数词 (Numerals)	一、5、千
	35. 数量词 (Quantifiers)	一些、两个、多次
	36. 量词 (Measure words)	这类、那种、各条
	37. 动量词 (Momentum quantifiers)	一番、两回、五场
	38. 时量词 (Time quantifiers)	两点、三周、十年
5 代词和代动词 Pronouns & Pro-Verbs	39. 第一人代词“我” (First person pronoun: I/我)	我
	40. 第一人代词“我们” (First person pronoun: we/我们)	我们
	41. 其他第一人称代词 (Other first person pronouns)	吾、咱们、晚生
	42. 第二人称代词 (Second person pronouns)	你、您、你们
	43. 第三人称代词 (Third person pronouns)	他、她、他们
	44. 代词: 它 (Pronoun: it/它)	它
	45. 普通名词+们 (Common nouns + plural/们)	朋友们、小鸟们
	46. 指示代词 (Demonstrative pronouns)	每、某、这儿
	47. 不定代词 (Indefinite pronouns)	某事、有人、任何
	48. 句首代词 (Pronouns at the beginning of a sentence)	我不信、你猜猜
	49. 代动词 (Pro-Verbs)	再来一杯、来个比赛
6 副词及其特殊属类 Adverbs & Special Adv.Genus	50. 副词: 最常用 (Adverbs: most commonly used)	也、不、就
	51. 副词: 中度常用 (Adverbs: moderately commonly used)	实在、不必、皆
	52. 副词: 非常用 (Adverbs: rarely used)	一概、一举、甚至于
	53. 副词: 表必然性 (Adverbs: inevitability)	甬、显然、势必
	54. 副词: 表可能 (Adverbs: possibility)	稍、大概、或许
	55. 副词: 表态度 (Adverbs: attitude)	难道、居然、不妨
7 介词及其短语 Prepositional Phrases	56. 句首介词 (Prepositions in the beginning of a sentence)	从、据、关于
	57. 介词短语 (Prepositional phrase)	向……、与……、由……

8	助词 Auxiliaries	58. 助词“的” (Auxiliary: de1/的)	的
		59. 助词“地” (Auxiliary: de2/地)	地
		60. 助词“得” (Auxiliary: de3/得)	得
		61. 助词“等”\“等等” (Auxiliaries: deng/等,dengdeng/等等)	等、等等
		62. 比况助词 (Situational comparative auxiliaries)	一样、似的
9	其他词汇属类 Other Lexical Genera	63. 语气词 (Modal particles)	啊、呢、焉
		64. 顺序词 (Sequential words)	首先、其次、总之
		65. 句内并列连词 (Syntactic conjunctions)	和、跟、以及
		66. 小品词 (Grammatical particles)	嗯、对对、行行
		67. 口语词 (Oral words)	今儿、够呛、没辙
		68. 儿化音 (Er hua words)	头儿、盖儿、手帕儿
		69. 嵌偶单音词 (monosyllabic words used in disyllabic templates)	极佳、围敌、返京
		70. 合偶双音词 (Couple disyllabic words)	禁止、购买、安装
		71. 古语词 (Classical words)	称之、岂不、颇有
		72. 增强语 (Amplification words)	深深、极其、颇为
		73. 模糊限制语 (Hedges)	至少、约摸、恐怕
10	名词形式 Noun Forms	74. 名物化 (Nominalization)	知名度、开放性
		75. 名动词 (Nominal verbs)	研究、储备、介绍
		76. 名形词 (Nominal adjectives)	方便、亲近、不同
11	状态形式 State Forms	77. 动词“是”作主要动词 (The verb 是/‘be’ used as the main verb)	尤其是、无疑是
		78. 动词“有”，表存现 (The verb 有/‘have’ used to express the existence)	有学者、有争议
12	时态和体标记 Temporal & Aspect Markers	79. 进行式动词 (Progressive verbs)	正看、玩耍中
		80. 过去式动词 (Past verbs)	曾听、刚刚来
		81. 助词“着” (Auxiliary: zhe/着)	着
		82. 助词“了” (Auxiliary: le/了)	了
		83. 助词“过” (Auxiliary: guo/过)	过
13	情态动词 Modal Verbs	84. 必要性情态动词 (Necessity modal verbs)	必须、一定、务必
		85. 情态动词：表未来 (Modal verbs: future)	会、要、应该
14	地点和时间状语 Adverbials of Place/Time	86. 时间副词 (Adverbs of time)	不断、偶尔、久久
		87. 副词：表处所 (Adverbs: Location)	到处、随处、遍地
		88. 处所词 (Location nouns)	内陆、海滨、前线
		89. 方位词 (Position words)	上、前、东南部
15	缩略形 Contractions	90. 缩略语 (Abbreviations)	北大、冬奥会
16	否定 Negatives	91. 否定词 (Negative words)	不、莫、未曾
17	独立语 Independent Words	92. 叹词 (Interjections)	咦、嗨、哎呀
		93. 拟声词 (Onomatopoeia)	嘭、咚咚、哗啦啦
		94. 插入语 (Parenthesis)	综上所述、相比之下
18	疑问句 Questions	95. 特指疑问句 (Interrogative Sentences)	哪些……、如何……
19	被动 Passives	96. 被动 (Passive)	被……、为……所……
		97. 无施事者被动句 (Passive sentences without doer)	被骂、挨打
		98. 介词“把” (Preposition 把/‘ba’ sentences)	把
20	复句 Compound Sentences	99. 并列复句 (Parallel compound sentences)	既……又……
		100. 顺承复句 (Successive compound sentences)	先……再……
		101. 解说复句 (Explanatory compound sentences)	即……、就是说……
		102. 选择复句 (Selective compound sentences)	或者……或者……
		103. 递进复句 (Progressive compound sentences)	不仅……而且……
		104. 条件复句 (Conditional compound sentences)	只要……就……
		105. 假设复句 (Hypothetical compound sentences)	如果……那么……
		106. 因果复句 (Causal compound sentences)	因为……所以……
		107. 目的复句 (Purpose compound sentences)	以便……、以免……
		108. 转折复句 (Turning compound sentences)	虽然……但是……
21	词汇丰度 Lexical specifications	109. 词汇多样性 (Lexical diversity)	/
		110. 平均词长 (Average word length)	/
		111. 平均句长 (Average sentence length)	/