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

2022 年度博士学位申請論文

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

大阪大学大学院言語文化研究科
言語文化専攻
徐　勤

目　次

第 1 章　序論 1

1.1 研究背景と目的 1

1.2 本研究の構成 2

1.3 使用データと分析ツール 3

1.3.1 使用データ 3

1.3.2 主なツール 5

1.4 主な分析手法 7

第 2 章　先行研究 11

2.1 多次元分析とは 11

2.2 英語の多次元分析 14

2.3 中国語の多次元分析 23

2.4 本研究の位置付け 24

第 3 章　言語項目の特定 27

3.1 言語項目 27

3.2 言語項目の特定 29

3.3 言語項目の集計 54

第 4 章　コーパスの構築 60

4.1 作文コーパスの構築 60

4.2 コーパスの検索機能とその応用 61

4.2.1 Keyword による Sentence の検索 61

4.2.2 品詞の検索 64

4.2.3 Wordlist の検索 69

4.2.4 Concordance の検索 72

4.3 本章のまとめ 75

第 5 章　多次元分析 76
5.1 コーパス 76
5.2 分析手順 76
5.3 リサーチクエスチョン 81
5.4 因子分析 82
5.4.1 共通性が低い変数の除去 82
5.4.2 KMO と Bartlett の球面性検定 85
5.4.3 因子数の決定 86
5.4.4 因子の負荷量 88
5.4.5 因子間の相関 90
5.4.6 各因子の寄与率 91
5.4.7 共起パターン 91
5.4.8 次元スコア 95
5.4.8.1 各テキストの次元スコア 95
5.4.8.2 各ジャンルの次元スコア 98
5.5 本章のまとめ 99

第 6 章　中国語作文における言語変異の次元 100
6.1 次元の解釈 101
6.1.1 次元 1：産出語彙の豊富さ VS. 産出語彙の限定性 104
6.1.1 次元 2：状態状況表現の適切さ 108
6.1.3 次元 3：副詞の修飾 111
6.1.4 次元 4：動作の描写 113
6.2 同一次元におけるジャンル間の比較 116
6.2.1 母語話者作文におけるジャンル間の比較 116
6.2.2 同一作文ジャンルにおける母語話者作文と学習者作文の比較 120
6.3 本章のまとめ 127

第 7 章　次元の言語項目におけるジャンル間の比較 129
7.1 次元 1 の言語項目 129
7.2 次元 2 の言語項目 138

7.3 次元 3 の言語項目 147
7.4 次元 4 の言語項目 150
7.5 本章のまとめ 155

第 8 章　中国語教育への応用 156
8.1 学習者の作文における特徴と問題点 156
8.2 在日中国語教育向けの提言 160
8.2.1 言語項目の過剰使用について 160
8.2.2 言語項目の過少使用について 178
8.2.3 中国語作文指導への提言 189
8.3 本章のまとめ 189

第 9 章　結び 190
9.1 本論のまとめ 190
9.2 本論の意義 191
9.3 今後の課題 192

参考文献 194

付録 I　NLPIR のタグセット 204

付録 II　言語項目 208

付録 III　HSK 単語リスト（計 5000 語） 212
HSK 1 級（150 語） 212
HSK 2 級（150 語） 213
HSK 3 級（300 語） 214
HSK 4 級（600 語） 216
HSK 5 級（1300 語） 219
HSK 6 級（2500 語） 225

謝辞 238

第 1 章　序論

1.1 研究背景と目的

外国語学習において習得すべき話す・聞く・書く・読むという四技能のうち、「書く」すなわち文章表現は、外国語教育において欠くことのできない学習分野である。輿水（2005: 150）は「書く」技能について、「ライティングとは、自分の考えや知っていることを書きことばで書きとめ他人に伝える技能をいう」と述べている。

学習者が実際に産出した書き言葉は、学習者の言語運用能力や言語使用に関して多くの特徴を教えてくれる（石川 2008: 202）。第二言語習得の研究では、学習とは第二言語（L2）の個々の語彙や文法の特徴を知ることだけでなく、これらの特徴がどのように連携して談話（discourse）の種類（例えば、物語や説明）を生成するのかを知ることであるとされている（Asención-Delaney & Collentine 2011）。Myles & Mitchell（2004）では、コーパス研究は、大量のデータを調べることによって、第二言語習得（SLA: Second language acquisition）研究の一般化可能性（generalizability）を補完し、高めることができると指摘している。とりわけこれに多次元分析（Multi-features / Multi-dimensional approach）を加えるならば、学習者により産出された言葉における言語特徴や言語変異（Linguistic Variation）の分析にも役立つことが知られている（Aguado-Jiménez *et al.* 2012）。

従来、上記に述べた四技能において、ライティングは難度が一番高いと言われてきた。近年、中国語学習者によって書かれた中国語作文に関する特徴に着目した研究（例えば、高 1996; 馬 2017）が多く行われてきたが、「日本語母語話者を対象とした作文教育に関する研究は依然として不足している」と楊（2022）が述べているように、既存の研究においては、日本人学習者が書いた中国語作文の特徴についての考察は十分になされていない。

そこで、本研究では、筆者が作成した中国語作文コーパスにもとづいて、日本人中国語学習者が書いた中国語作文の言語特徴を、中国語母語話者の作文と対照しつつ多次元分析法によって明らかにし、そこから日本語母語話者に対する中国語教育の問題点を探り、中国語の作文指導や教材の改善に資することを目的とする。

1.2 本研究の構成

本研究は全 9 章から構成される。

第 1 章「序論」では、本研究の背景と目的を明らかにした上で、本論文の構成を示す。そして、本研究で使用するデータとツール、及び主な分析手法について紹介する。

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。第 1 節では、多次元分析の定義を説明する。第 2〜3 節では、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。第 4 節では、本研究の位置付けを示す。

第 3 章「言語項目の特定」では、変数として多次元分析に用いる言語項目（Linguistic Features）を決める。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。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。

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使用し、学習者の作文コーパスを構築する。第 1 節では、作文テキストの収集・分類及びコーパスの内訳を明記する。第 2 節では、コーパスの検索機能、キーワードによるセンテンスの検索（Keywords in Sentence）、品詞（Part-of-Speech）・ワードリスト（Wordlist）・コンコーダンス（Concordance）の検索など、及び実際に検索した場合の手順と結果例を示す。

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を説明する。

第 6 章「中国語作文における言語変異の次元」では、第 5 章で得られた因子（言語項目の共起パターン）を主な次元として捉え、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。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い、同一ジャンルにおける学習者作文と母語話者作文間の言語変異を調べる。

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を調べる。また、言語項目の出現頻度（相対頻度）を用いて多重比較を行うことで、学習者の作文における言語項目の使用傾向や問題点を考察する。

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る。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に関する特徴及び問題点の原因を解明する。最後に、中国語教育内容向けの提言を行う。

第 9 章「結び」では、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言及する。

1.3 使用データと分析ツール

1.3.1 使用データ

（1）作文コーパスについて

分析対象の作文データは、著者編纂の中国語作文コーパス（2022 年 9 月に一般公開されている、URL：https://xuqin-composition-corpus.herokuapp.com/）に収録された作文テキストである。このコーパスは、日本人中国語学習者と中国語母語話者が書いた中国語作文によって構築されたものである。

コーパスにおける日本人中国語学習者の作文は、2020 年から 2022 年にかけて、大阪大学外国語学部中国語専攻に在籍しており、「中級中国語作文」授業を受講している大学 2 年生（計 230 名）と、「上級中国語作文」授業を受講している大学 3 年生（計 105 名）の日本人中国語学習者によって書かれた作文を収集して電子化したテキストである。「中級中国語作文」授業では、主に手紙、日記、人物描写、出来事の叙述などのような作文ジャンルを中心に学習者に作文を書かせた。「上級中国語作文」の授業では、人物描写、出来事の叙述の練習に加え、意見文の書き方の練習も始まった。これで、学習者が書いた 5 つの作文ジャンル（手紙・日記・出来事の叙述・人物描写・意見文）からなる約 700 の作文テキストを収集した。

作文コーパスを構築する段階では、収集した作文を、学習者の中国語レベルと受講するコースによって、中級作文コーパスと上級作文コーパスに分類し、学習者の作文コーパスを構築した。さらに、 学習者の作文における言語特徴と比較対照するために、中国語母語話者が書いた同じジャンルの作文を収集する必要がある。張(2018)[1]では、高校 3 年生の中国語レベルは中国語母語話者の言語レベルを代表することができると述べてい

[1] 張(2018)：“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”。

るため、本研究は“zuowen 网” （中国語作文の WEB サイト[2]）に掲載される高校 3 年生の中国人が書いた作文（ジャンル：手紙・日記・出来事の叙述・人物描写・意見文）を抽出し参照コーパスを構築した。

上記に述べたように、作文コーパスには、中国語学習者と中国語母語話者によって書かれた 5 つの作文ジャンル[3]からなる作文テキストが含まれている。本研究の各ジャンルにおける作文の特徴について説明する。手紙（中国語では“书信”）は主に友人やクラスメート、家族の人、先生などに手紙を書くこと、日記（“日记”）はその日の出来事を記録すること、出来事の叙述（“叙事”）はある出来事を文字にして語ること、人物描写（“人物描写”）はある人物を外見、話し方、行動、心理などの面から説明すること、意見文（“议论文”）は事実や根拠を示しながら、自分の意見や考えを直接的に表現することである。コーパスの構築および設けた検索機能の詳細については、第 4 章を参照されたい。

(2) 作文例文の出典について

本論で取り扱う中国語学習者の例文は、中国語を専攻としている日本人大学生が書いた作文の抜粋であり、中国語母語話者の例文は、高校三年生が書いた作文からの引用である。また、第 7 章では、中級中国語作文と上級中国語作文の授業で使用される教科書における例文も用いる。著作権と研究倫理の関係上、例文を引用する際、例文の文末に

---

[2] “zuowen 网”（中国語作文 WEB サイト）には、小学校 1 年生から高校 3 年生までの学生が書いた様々なジャンルの作文が掲載されている。URL：http://www.zuowen.com/gaozhong/〔最終アクセス日: 2022 年 5 月 20 日〕

[3] 中国語の作文において、「手紙」（例えば：第 1 章の 9 ページにある例文）と「日記」（例えば、第 7 章にある例文 26～27）は、構成や体裁が慣習的に決まっているジャンルである。中国語を専攻とする大学 2 年生の作文授業では、中国語の手紙の書き方を習得するために手紙を書くことを練習し、同時に物語文（出来事の叙述）の基礎を作るために、日記を書く練習もする。「日記」、「手紙」、「人物描写」と「出来事の叙述」という 4 つのジャンルは、それぞれ絶対的に独立したジャンルというわけではない。なぜなら、これら 4 種類のジャンルは、いずれも出来事に対する記述、人物に対する描写、情感の表出などをその内容に含めることができるからである。とはいえ「日記」や「手紙」には決まった形式があるし、「人物描写」と「出来事の叙述」では記述の重点が異なり、前者が（人物の外見やある出来事の記述を通して）人物の特徴を際立たせることに重きを置く一方、後者は出来事そのものの記述に主眼がある。第 6 章の図 6.5 に示すように、これら 4 種類のジャンルにおいては、4 つの次元の全体的分布にある程度の類似性がみられ、絶対的に異なるわけでも、さりとてまったく同一でもないのは、このように各ジャンル間の作文内容に一定の重複が見られることに起因している。

JP2 （中級中国語学習者）、JP3（上級中国語学習者）、CHN（中国語母語話者）、JP2_textbook（中級中国語作文授業の教科書）、JP3_textbook（上級中国語作文授業の教科書）のように注記して引用元を明らかにする。

1.3.2 主なツール

分析用の主なツールは Python、Streamlit、Heroku、NLPIR などである。

Python は統計処理、WEB アプリ開発、機械学習などで利用されるオープンソースのプログラミング言語である。Streamlit とは Python を用いて Web アプリを容易に作成できるフレームワーク（Framework）であり、Streamlit ホームページ「URL：https://docs.streamlit.io/library/api-reference（最終アクセス日：2022 年 11 月 15 日）」には様々な API 機能が載っている。Streamlit で作成した Web サイトやアプリは、Github 経由で Heroku と連携し、Heroku にデプロイされると公開することができる（注：Heroku の無料サービスが 2022 年 11 月 28 日で終了となった）。

NLPIR（中国語の全称：“NLPIR-ICTCLAS 汉语分词系统”）[4]は、中国語の文章を単語に分かち書きして、それぞれの単語の品詞タグを自動的に付与できるシステムである。このツールは Huaping Zhang 氏によって開発された中国語の単語分割と品詞解析のシステムであり、Big data 検索及び data mining 処理のプラットフォーム（中国語：大数据搜索与挖掘平台）である。NLPIR のタグセット（Tagset）には、記号と句読点以外に 17 種類の品詞タグが設定されている。NLPIR のタグセットにおけるタグの分類は、本稿末の付録 I を参照されたい。

黄・廖（2017 下冊: 8-9）は中国語の品詞を内容語と機能語に分類している。内容語は名詞、動詞、形容詞、区別詞、数詞、量詞、副詞、代名詞、及び特殊内容語の擬音語・感嘆詞に細分化され、機能語は前置詞、接続詞、助動詞、語気詞に細分化されていると述べられている（黄・廖 2017 下冊：8-9）[5]。また、名詞には、（1）固有名詞、（2）普通名詞、（3）時間名詞、（4）場所名詞、（5）方位名詞があると黄・廖（2017 下冊：9）

[4] NLPIR（中国語の全称：“NLPIR-ICTCLAS 汉语分词系统”）：http://ictclas.nlpir.org/〔最終アクセス日：2022 年 6 月 5 日〕

[5] 黄・廖（2017 下冊：8-9）：“实词细分为名词、动词、形容词、区别词、数词、量词、副词、代词，以及特殊实词拟声词、叹词；虚词细分为介词、连词、助词、语气词”。

が述べている[6]。表 1.1 は黄・廖（ 2017）と NLPIR の品詞分類を整理して示したものである。

表 1.1　黄・廖（ 2017 下冊: 8-9）と NLPIR の品詞分類

《现代汉语》（現代中国語） 黄・廖（2017: 7-35）		NLPIR Tagset の品詞分類[7]	
14 種類の品詞		17 種類の品詞	タグ
内容語 （实词 / content words）	名詞（名词）	名詞　（名词）	/n
		時間詞（时间词）	/t
		場所詞（处所词）	/s
		方位詞（方位词）	/f
	動詞（动词）		/v
	形容詞（形容词）		/a
	区別詞（区别词）		/b
	数詞（数词）		/m
	量詞（量词）		/q
	副詞（副词）		/d
	代詞（代词）		/r
	擬音語（拟声词）		/o
	感嘆詞（叹词）		/e
機能語 （虚词 / function words）	前置詞（介词）		/p
	接続詞（连词）		/c
	助動詞（助词）		/u
	語気詞（语气词）		/y

注：カッコ内は中国語の名称を示している。

本研究では、NLPIR を用いて中国語の単語分割とタグつけを行った上で、Python で各テキストに対して言語項目の頻度を集計して因子分析を行う[8]。また、Streamlit の API 機能に基づいてオンライン作文コーパスを構築し、Heroku と連携して作文コーパスを公開する。

---

[6] 黄・廖（2017 下冊: 9）：“名词有以下几种：（1）专有名词；（2）普通名词；（3）时间名词；（4）处所名词；（5）方位名词”。

[7] NLPIR Tagset（中国語名：“计算所汉语词性标记集”）：
http://kgb.lingjoin.com/nlpir/html/readme.htm〔最終アクセス日：2022 年 5 月 5 日〕

[8] 言語項目の頻度集計に用いられたツールは、筆者が Python を用いて作成した頻度集計のプログラムである。因子分析を行うツールとしては、「factor_analyzer」という Python のパッケージを用いた。

1.4 主な分析手法

(1) 多次元分析法

本研究の分析に用いられる分析手法は多次元分析法(Multi-features / Multi-dimensional approach)である。コーパスに基づく多次元分析は、レジスター(register)やジャンル(genres)の違いが言語共起の基礎的パターンに関係するという理論的な仮定に基づき、Douglas Biber(例、Biber 1988, 1995, 2006 など)によって開拓されたものである。例えば、作家があるレジスターから別のレジスターに移るとき、特定のメッセージを伝え、特定の聴衆とつながるために、共起する言語特徴(co-occurring features)のあるセットから別のセットの共起する言語特徴へと自然に移行するのである(Hardy & Friginal 2012)。

上に述べたジャンル(genres)とレジスター(register)は隣接する概念であるとはいえ、この 2 つの用語は「最も紛らわしいもので、主にある程度重複しているため、しばしば同じ意味で使われることがある」と Lee(2001)[9]は述べている。Biber & Conrad(2019: 15)[10]によれば、ジャンルとレジスターは異なる種類のテキストではなく、テキストの多様性を分析するための異なるアプローチまたは見方とみなされており、同じテキストをレジスター、ジャンルの見方で分析することが可能である。

ジャンルとレジスターの定義について、Biber & Conrad(2019: 16)のレジスター、ジャンルに関する特徴の分類により引用すると下記(表 1.2)のようになる。

[9] 'The two terms genre and register are the most confusing, and are often used interchangeably, mainly because they overlap to some degree' (Lee 2001).

[10] Biber & Conrad (2019: 15): '[W]e regard genre, register, and style as different approaches or perspectives for analyzing text varieties, and not as different kinds of texts or different varieties. In fact, the same texts can be analyzed from register, genre, and style perspectives'.

表 1.2　レジスター、ジャンルの特徴に関する定義

Defining characteristic	Register	Genre
textual focus	sample of text excerpts	complete texts
Linguistic characteristics	any lexico-grammatical feature	Specialized expressions, rhetorical organization, formatting
distribution of linguistic characteristics	frequent and pervasive in texts from the variety	usually once-occurring in the text, in a particular place in the text
interpretation	features serve important communicative functions in the register	Features are conventionally associated with the genre: the expected format, but often not functional

注：Biber & Conrad（2019: 16）の ‘Defining Characteristics of Registers, Genres, and Styles’ に基づいて作成

レジスター（register）は言語使用域とも呼ばれ、言語の用途に応じた変化（すなわち機能的な多様性）を指す言葉として使われ（Aguado-Jiménez *et al.* 2012）、特定の言語使用を特徴づける語彙的および文法的特徴の組み合わせである（Halliday & Hasan 1989; Martin 1992）。レジスターの違いは、単語やフレーズの選択、および節の構成や連結の仕方に現れる。特定のレジスターを構成する言語特徴の構成は、人間がテキストを聞いたり読んだりする際に、そのテキストがどのような文脈で、どのような目的のために作られたかという印象を形成することが可能である（Schleppegrell 2001）。Aguado-Jiménez *et al*（2012）によると、同じ機能を持ち、同じ読者層を対象とするテキストは、ほとんどの場合、同じレジスターに属し、類似した言語特徴を共有する可能性が高い。例えば、ある特定のレジスター（例、手紙）に頻繁に現れる特徴（例、手紙における人称代名詞 I や You など）は、別のレジスター（例、公文書）では通常稀である。

ジャンル（genres）とは、特定の文化的な文脈で達成される、目的を持った、段階的な言語の使用である（Christie 1985）と述べられている。また、Biber（1988: 170）[11]は、「ジャンルの分類は話し手の目的や話題に関連する外部基準に基づいて決定され、形式ではなく、使用に基づいて割り当てられる」と主張している。

表 1.2 から分かるように、レジスター的な見方では、多様なテキストの中から代表的なものを抜粋して広範な語彙・文法的な特徴に焦点が当てられるのに対し、ジャンル的

[11] ‘Genre categories are determined on the basis of external criteria relating to the speaker's purpose and topic; they are assigned on the basis of use rather than on the basis of form’（Biber 1988: 170）.

な見方では、テキスト全体を構成するために使用される言語特徴に注目する（Biber & Conrad 2019: 15）。中国語の手紙という体裁を例として挙げる。例えば、下記の例文に示したように、中国語の手紙では、最初に呼びかけの言葉を書き、呼びかけの後ろにコロン「：」が入り、その後挨拶（例えば：“你好”）などを書いて本文が入り、本文が終わったら挨拶（例えば：“祝你……”）を書き、最後に右下に差出人の署名と日付が記される。このように、手紙全体の構成や体裁が慣習的に決まっている。

教科書における例文：《给新朋友的信》「友達への手紙」

蔡云小姐：

你好！迟寄为歉。

因 2 月 28 日之前从昆明至北京的火车票、飞机票都买不到，我不得不坐 3 月 1 日下午的飞机回北京，第二天早晨 8 点就开始了新学期的课。一转眼半个月已匆匆忙忙地过去了，直到现在才给你寄信，真对不起。

我每天看着你给我的书包，上面绣着“各国人民心相连”的字样，而且是你把“全”字改成了“各”。我觉得北京和景洪、日本和中国人民的心连接了起来。36 天的云南之行，在西双版纳的日子给我留下了最深刻的印象，特别是你和你的父母，还有那只可爱的小狗。为了纪念，我把那时拍的照片寄给你，请笑纳。

若有机会来北京，请你一定跟我联系。希望能进一步加深我们的友谊。再见！

祝你

身体健康、工作顺利！

川上靖雄

×年×月×日

注：亓华・贾丹丹(2013) の『汉语・纵横 写作课本(上)』p.26 に掲載されている例文より引用

Biber & Conrad(2019: 17)によれば、レジスター的な見方とジャンル的な見方は、適用できる範囲が異なる。レジスター的な見方に関連する言語特徴を特定するためには、テキスト全体が必要であり、テキストの抜粋は、ジャンルを定義する言語的慣習（例えば、テキストの始まりや終わりの慣習的な形式）を必ずしも表していないため、ジャンル分

析には適切ではない（Biber & Conrad 2019: 17）[12]と述べられている。従って、本研究の1.3.1 節では、作文の体裁（例、手紙・日記・出来事の叙述など）を指す言葉として、「ジャンル」という用語を用いている。

(2) 因子分析

多次元分析の常として、因子分析で次元を特定し、その次元を使ってサブコーパスを比較する（Thompson *et al.* 2017）[13]と述べられている。因子分析は、言語の共起パターンを分析し、大きな変数（個々の言語項目）の集合を、より小さな言語的パラメータの集合に還元する多変量解析の手法である。社会科学の研究において、因子分析は、データ削減のために最もよく使われる方法の一つである。この統計手法は基礎となる次元または因子が複雑な現象を説明するために使用することができると仮定しており、その目的は、観察可能または測定可能な指標（変数）の大きいセットに基づいて、直接観察可能でない因子を識別することである（Lingard & Rowlinson 2006）と述べられている。

テキスト研究における因子分析の最も早い使用例は、Carroll（1960）によるものである。Carroll（1960）は因子分析法の技術を用いて、39 種の客観的な言語測定尺度（linguistic measures）と 29 種の主観的な知覚的評価（perceptual ratings）のセットを、言語的スタイル（linguistic style）の基本パラメータと解釈して、6 つの因子に特定した。レジスター間の類似点と相違点は、これらの言語パラメータ（因子）のそれぞれに沿って記述される。まとめて言うと、言語使用域やレジスター（register）に関する言語変異の研究分野では、因子分析は多次元分析と呼ばれる研究手法で最もよく適用されている（Egbert & Biber 2018）。

---

[12] Biber & Conrad (2019: 17): ‘The register and genre perspectives differ in the extent to which they can be applied. Complete texts are required to identify the linguistic characteristics associated with the genre perspective. Text excerpts are not adequate for genre analysis, because they do not necessarily represent the linguistic conventions that define the genre (e.g, the conventional ways to begin or end a text)’.

[13] ‘As is usual in MDA, factor analysis is used to identify dimensions, and the dimensions are used to compare sub-corpora’ (Thompson et al. 2017).

第 2 章　先行研究

本章では、多次元分析に関する先行研究について概観する。まず、第 1 節では、多次元分析の定義を明らかにする。そして、第 2〜3 節では、英語と中国語への多次元分析の応用に関する先行研究を整理する。最後に、第 4 節では、先行研究から得られた知見によって、本研究の位置付けを示す。

2.1 多次元分析とは

テキスト分析の研究分野において、ジャンルの変異（genre variation）を調査するための多次元分析法（MF/MD 法：Multi-Dimensional approach、Multi-features / Multi-dimensional approach とも呼ばれる）は、Biber（1986）で初めて概説され、Biber（1988）が発展させたものであり、話し言葉と書き言葉のレジスター（register）間の言語変異のパターンを調べ、テキスト全体の言語パターンを記述する目的で特別に開発された統計的手法である（例、Biber 1986, 1988, 1995）。Douglas（1992）によれば、多次元分析法には 8 つの顕著な特徴がある。

Douglas（1992: 332）：

① It is corpus based, depending on analysis of a large number of naturally occurring texts.

② It is computer-based in that it depends on automated analyses of linguistic features in texts. This characteristic enables distributional analysis of many linguistic features across many texts and text varieties.

③ The research goal of the approach is the linguistic analysis of texts, genres, text types, styles or registers, rather than analysis of individual linguistic constructions.

④ The importance of variationist and comparative perspectives is assumed by the approach. That is, the approach is based on the assumption that different kinds of text differ linguistically and functionally, so that analysis of any one or two text varieties is not adequate for conclusions concerning a discourse domain (e.g., speech and writing in English).

⑤ The approach is explicitly multi-dimensional. That is, it is assumed that multiple parameters of variation will be operative in any discourse domain.

⑥ It is quantitative. Analyses are based on frequency counts of linguistic features, and multivariate statistical techniques are used to analyze the relations among linguistic features and among texts.

⑦It synthesizes quantitative and functional approaches. That is, the statistical analyses are interpreted in functional terms, to determine the underlying communicative functions associated with each distributional pattern. The approach is based on the assumption that statistical co-occurrence patterns reflect underlying shared communicative functions.

⑧ It synthesizes macroscopic and microscopic approaches. That is, macroscopic investigations of the overall parameters of linguistic variation, which are based on analysis of the distribution of many linguistic features across many texts and genres, are complemented by detailed analyses of particular features in particular texts.

上記の引用における①～②に示したように、テキストの分析およびテキストにおける言語特徴の自動分析に依存するという点で、多次元分析法はコーパスベース（corpus based）であり、コンピュータベース（computer-based）である。 石川（2012: 13）によれば、コーパス（corpus）とは、「（1）書き言葉や話し言葉などの現実の言語を、（2）大規模に、（3）基準に沿って網羅的・代表的に収集し、（4）コンピュータ上で処理できるデータとして保存し、（5）言語研究に使用するもの」である。また、コーパスに基づく分析の特徴に関して、Biber *et al*（1998: 4）のコーパスに基づく分析の本質的な 4 つの特徴を、斉藤ら（2003: 10）の訳により引用すると下記のようになる。

Biber *et al*（1998: 4）（斉藤ら訳 2003: 10）：『コーパス言語学―言語構造と用法の研究』南雲堂

1) それは実証的であり、自然なテキストにおける用法の実際のパターンを分析する。

2) それは分析の基礎として、「コーパス」として知られている、自然なテキストの大きくて一定の原則に基づいて収集されたものを利用する。

3) それは分析にコンピュータを広範に利用し、自動的と双方向的な手法の両方を活用する。

4) それは数量的と質的の両方の分析手法に依存する。

コンピュータの「分析中に気が変わったり、疲れたりすることはない」（斉藤ら訳 2003: 11）という利点で、手作業で扱えない膨大なコーパスデータやデータベースを機械によって蓄積し、読み取って分析することができる。この特長により、多くのテキストにまたがる様々な言語特徴の分布を分析することが可能になる。

上記の Douglas（1992: 332）の引用における③～⑤から分かるように、多次元分析は、個々の言語構文の分析ではなく、テキスト・ジャンル・スタイル・レジスターなどの談話領域（discourse domain）に関する言語分析に着目し、変異論（variationist）的・比較論（comparative）的な視点が重要である。

この分析手法は、異なるタイプの談話領域は言語的・機能的に異なるという前提に立ち、複数の変異パラメータが作用していることを想定している。この想定に基づき、引用の⑥～⑧から明らかなように、多次元分析は量的な多変量統計技法（multivariate statistical techniques）によって言語項目及びテキスト間の関係を分析し、統計的に共起しているパターンを特定する。そして、各分布パターンに関連する根本的な伝達機能（communicative functions）を機能的な用語で解釈し、次元として捉える。Egbert & Biber（2018）によると、多次元分析法に関わる次元には 3 つの大きな特徴がある[14]。

Egbert & Biber（2018）：

a) 各次元は、共起する言語特徴の明確な集合によって統計的に定義される。

b) 各次元に関連する言語使用域の変異のパターンが異なる。

c) 各次元は特定の伝達機能（communicative functions）と関連している。

共起パターン（次元）が、根底にある共通の伝達機能を反映しているという仮定に基づいている。例えば、Biber（1988）は、著書『Variation Across Speech and Writing』で共起する言語項目（co-occurring linguistic features）の概念を紹介し、共起する言語項目のクラスター（clusters）を特定し解釈する方法論として多次元分析法を提唱した。Biber（1988）によれば、あらゆるレジスター（registe）を研究・分析するために最も必要な条件は、ある言語項目同士がクラスターとして共起し、そのレジスターの特徴（features）

[14] “There are three major characteristics of MDA: (a) Each dimension is defined statistically by a distinct set of co-occurring linguistic features; (b) There are different patterns of register variation associated with each dimension; (c) Each dimension is associated with particular communicative functions.” (Egbert & Biber 2018).

を反映することであると考えられる。Ahmad & Mahmood（2015）が述べているように、「多次元分析は、レジスター変異の研究において、言語項目の共起に重点を置き、個々の言語項目ではレジスターを確実に区別することができないことを強調している」[15]。

多次元分析は主に、統計的因子分析（factor analysis）によって共起する言語項目の集合を特定し、言語項目間（変数間）の共起関係（co-occurrence）を明らかにして言語変異を多次元的に捉える。次元が確立された後、次元と呼ばれる共通伝達機能（communicative functions）の観点から解釈し、言語項目によって示される共通伝達機能に基づいて解釈ラベルを付ける。

因子分析（factor analysis）は多次元分析と呼ばれる研究手法で最もよく適用されている（Egbert & Biber 2018）。「因子分析を使えば、言語項目のうちのどれがテキスト内で共起しやすいかを示すことができる」（斉藤ら訳 2003: 160）と述べているように、因子分析を使用することで、大量の言語項目は少数の因子に圧縮され、テキスト内で言語項目間の共起パターン（co-occurrence patterns）を特定できる。特定された共起パターンは、その根底にある機能的な関連性に基づいて、言語的な「次元」として解釈される。

従って、統計的因子分析に基づいた多次元分析手法は、「主観的な意味カテゴリーに基づくのではなく、主に客観的な文法的・語彙的特徴に基づいて言語変異を研究するボトムアップ・アプローチ（bottom-up approach）である」[16]（Egbert 2012）といえる。

2.2 英語の多次元分析

（1）書き言葉と話し言葉の言語変異への応用

話し言葉と書き言葉のレジスター間の相違は、言語学者には興味深い課題である。英語では、話し言葉と書き言葉のレジスターはどのように違うのかという課題に対して、多次元分析と呼ばれるレジスター変異の研究が多く見られる。英語の書き言葉と話し言葉の多次元分析において、Douglas Biber の研究成果（Biber 1986, 1988, 2006; Biber *et al.* 2002 等）は注目に値する。

---

[15] “The multi-dimensional (MD) approach lays emphasis on the co-occurrence of linguistic features in register variation studies and highlights the fact that individual linguistic features cannot reliably distinguish among registers” (Ahmad & Mahmood 2015).

[16] “MD analysis is a bottom-up approach to researching linguistic variation based on mostly objective grammatical and lexico-grammatical features rather than subjective semantic categories.” (Egbert 2012).

Biber（1986）は、英語の話し言葉と書き言葉のテキストタイプ間の類似点と相違点を説明するために、多次元分析のアプローチを用いた。16 種のレジスターに跨る 545 個のテキストサンプル（各テキストサンプルの語数：約 2000 語）を対象に、41 個の言語項目の分布に関する因子分析（多変量統計手法）を行なった結果、3 つの基本的な言語変異の次元が特定された。

Dimension 1: Interactive vs. Edited Text

（次元 1：インタラクティブ vs. 編集されたテキスト）[17]

Dimension 2: Abstract vs. Situated Content

（次元 2：抽象的 vs. 状況的な内容）

Dimension 3: Reported vs. Immediate Style

（次元 3：報告的 vs. 即時的スタイル）

Biber（1986）の研究は話し言葉と書き言葉における 16 種類のジャンルの類似点と相違点を明らかにしただけでなく、テキスト分析における多次元分析の重要な価値を実証した。そして、Biber（1988）は、英語の話し言葉コーパス（London-Lund Corpus）と書き言葉コーパス（LOB Corpus）から 21 ジャンル（例、新聞報道・論説・伝記・宗教・政府刊行物・学術散文・推理小説・ユーモア・電話による会話・即席スピーチ・準備されたスピーチ等）に Biber の収集した「私信」と「ビジネスレター」を加えた 23 ジャンルのテキストをデータとして用いて、ジャンル間の言語変異を考察した。67 種の言語項目（例、一人称代名詞・単語の長さ・疑問文・形容詞・副詞・法助動詞・等位接続詞等）の出現頻度を集計して相対頻度（1,000 語あたり）に換算し、因子分析を行った。その結果、共起する言語項目群の分布パターンを 7 つに特定し、その 7 つのパターンを主な次元（Dimension）として捉えた。さらに、各次元に共起する言語項目の共通伝達機能を吟味し、7 つの次元に下記のような機能ラベルを付与している。

Dimension 1: Involved versus Informational Production

（次元 1：関与性 vs. 情報提示的言語生成）

[17] 括弧内は筆者による次元名の和訳である。

Dimension 2: Narrative versus Non-narrative Concerns

（次元 2：物語調 vs. 非物語調ディスコース）

Dimension 3: Explicit versus Situation-Dependent Reference

（次元 3：綿密な指示 vs. 状況依存的指示）

Dimension 4: Overt Expression of Persuasion

（次元 4：論議の明示度）

Dimension 5: Abstract versus Non-Abstract Information

（次元 5：抽象的 vs. 非抽象的情報）

Dimension 6: On-Line Informational Elaboration

（次元 6：即席の情報提示の綿密性）

Dimension 7: Academic Hedging

（次元 7：学術的婉曲性）

Biber（1988）は上記に示した 7 つの次元を統計的な因子分析によって抽出したが、次元 6 と次元 7 は言語項目数が少なく、統計的な有意性が非常に低いため実際の分析には使用されておらず、最終的に最も有意性の高い 5 次元（次元 1～5）に絞って分析を行なっている。Biber（1988）は、多次元分析を、コーパスに基づく言語変異の定量的研究、実証的（empirical）研究、比較（comparative）に基づく研究に最も適した代替アプローチであるとみなしている。

Biber *et al*（2002）は「TOEFL 2000 Spoken and Written Academic Language Corpus」におけるレジスターの変異（register variation）を多次元分析手法で調査した。その分析により、大学のコンテクスト（university contexts）において、話し言葉のレジスターは書き言葉のレジスターと基本的に異なることが分かった。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、言語変異を解釈するための 4 つの次元を特定した。

Dimension 1: Oral vs. literate discourse

（次元 1：口語的談話 vs. 文語的談話）

Dimension 2: Procedural vs. content-focused discourse

（次元 2：手続き型談話 vs. 内容重視型談話）

Dimension 3: Reconstructed account of events

（次元 3：再構成された出来事の説明）

Dimension 4: Teacher-centered stance

（次元 4：教師中心のスタンス）

上に述べたように、「コーパスに基づく研究の枠組みを使って多次元分析を行えば、話し言葉と書き言葉のレジスターにわたる言語項目の主要な変異パターンを調べることができる」（斎藤ら訳 2003: 151）。Biber の多次元分析モデルは当初、上記のような互いに大きく異なると予想されたレジスターを区別するために使用されるが、言語変異の研究分野に大きく貢献しており、広範囲に応用されてきた。例えば、学術分野や学問領域（例、Biber *et al.* 1998; Kanoksilapatham 2007; Gray 2013; Egbert 2016）、異なる学問分野における学生のライティング（例、Hardy & Römer 2013; Gardner *et al*. 2015; Hardy & Friginal 2016）、L2 ライティング（例、Friginal & Weigle 2014; Asención-Delaney & Collentine 2011）などの言語変異の研究では、Biber の多次元分析と同じ方法論を用いて、言語変異を解釈する新しい次元を特定することができた。下記では、多次元分析に関する応用例の一部を整理する。

(2) 作家や文学作品間の言語変異への応用

Biber & Conrad（2019: 132-138）は、文法的特徴の使用パターンにより、作家や文学作品間の顕著な差異が明らかになった。Egbert（2012）では、多次元分析の応用により 19 世紀末のフィクション作家の文体（the writing styles of late 19th century fiction authors）を、言語的変数の共起パターンでどの程度区別できるかを調べた結果、文体の変異の根底にある 3 つの次元を特定しており、その 3 つの次元はそれぞれ「Thought Presentation versus Description（思考提示 *vs.* 描写）」、「Abstract Exposition versus Concrete Action（抽象的説明 *vs.* 具体的行動）」、「Dialogue versus Narrative（対話 *vs.* 物語）」として解釈された。

(3) ウェブやメディアの言語変異への応用

インターネットが人間のコミュニケーションのプラットフォームとして発展するにつれ、現代のネット上のレジスターの言語変異に関する研究に焦点を当ててきた。

例えば、Grieve *et al*（2010）は多次元分析の研究視点に基づき、200 万語の英語ブログ（blog）コーパスにおける言語変異（linguistic variation）を考察した。因子分析を行った結果、言語変異を反映できる 4 つの主要な次元が特定された。この 4 つの次元はそれぞれ、「Informational vs. Personal Focus（情報的焦点 vs. 個人的焦点）」、「Addressee Foucus（宛先焦点）」、「Thematic Variation（主題的変化）」、「Narrative Style（物語的スタイル）」であり、ブログというレジスターの機能的言語変異（functional linguistic variation）の重要なパターンを示している。

Hardy & Friginal（2012）は、Grieve *et al*（2010）によって特定された言語変異の 4 つの次元を用い、フィリピン人とアメリカ人の著者によって英語で書かれたインターネットブログとオンラインオピニオンコラム（online opinion columns）の 4 つのコーパスから、異文化間（cross-cultural）およびレジスター間（cross-register）の差異を調査した。その結果、フィリピン人とアメリカ人のテキストの言語構成は、4 つの機能的次元、すなわち、1）情報的焦点 vs. 個人的焦点、2）宛先焦点、3）主題的変化、4）物語的スタイルに沿って、著しい変化があることを示しており、アメリカのブログはフィリピンのブログに比べて、より個人的で、宛先に焦点を当てた傾向があるなど、大きな差異があることが明らかになった（Hardy & Friginal 2012）。

Biber & Egbert（2016）は多次元分析を用いて、検索可能なウェブ（searchable web）上の言語変異（linguistic variation）の次元と、それらの次元に関するウェブレジスター（web registers）間の類似点と相違点を探った。Sardinha（2018）は、オンライン・コミュニケーションにおける特定のレジスター（メール、ブログ、ウェブページ、Facebook 投稿、ツイート）が多次元分析の枠組みにおいてどの程度変異するかを検証し、レジスター間の実際の言語変異と類似性を考察した。その結果、インターネットレジスター間の言語変異を反映できる 3 つの次元、すなわち、「involved, interactive discourse versus informational focus（関与する対話的談話 vs. 情報焦点）」、「expression of stance: interactional evidentiality（スタンスの表現：対話的証拠性）」、「expression of stance: interactional affect（スタンスの表現：相互作用的影響）」が得られた（Sardinha 2018）。

Ahmad & Mahmood（2015）は多次元分析を通じてパキスタンの印刷メディアにおける報道記事（press reportage）の言語使用域を考察した。パキスタンの報道記事をイギリスの報道記事と比較した結果、パキスタンの報道記事は、情報量・物語性・明示性・非抽象性が高いのに対し、説得・論証の表現が最もあからさまでないことが明らかになった（Ahmad & Mahmood 2015）。

また、China Daily（CD）と The New York Times（NYT）の社説の書き方を比較するために、Huang & Ren（2020）は多次元分析を行い、5 つの次元を抽出し、次元 1 から次元 5 を「Involved vs. Informational Production（関与性 VS. 情報提示的言語生成）」、「Narrative vs. Non-Narrative Concerns（物語調 VS. 非物語調ディスコース）」、「Expression of Institutional Stance（組織的なスタンスの表明）」、「Overt Expression of Argumentation（主張の明示度）」、「Abstract Style（抽象的なスタイル）」のようにラベルを付けた。社説の書き方を比較した結果、CD 社説（China Daily）は NYT 社説（The New York Times）に比べ、より情報量が多く、非人称的（impersonal）で、より組織的な発言力を示すのに対し、NYT 社説は CD 社説よりも関与が強く、物語的で、あからさまに議論的で、個人的スタンスを示す傾向があることを明らかにした（Huang & Ren 2020）。Huang & Ren（2020）の研究は、米中両国の英語による説得的な文章に見られる独特なスタイルについて、新しい知見を提供するものであると言える。

(4) 学術や学問領域の言語変異への応用

近年は学問分野間の言語変異（Linguistic variation across disciplines）に関する研究も顕著である。例えば、Hardy & Römer（2013）は、MICUSP コーパスに含まれる学問分野（disciplines）を区別する 4 つの次元（① Involved, Academic Narrative versus Descriptive, Informational Discourse; ② Expression of Opinions and Mental Processes; ③ Situation-Dependent, Non-Procedural Evaluation versus Procedural Discourse; ④ Production of Possibility Statement and Argumentation）を特定し、分野別・ジャンル別の類似点と相違点、そして分野別・レジスター別教育法への示唆を考察している。Gray（2013）は多次元分析法を用い、学術分野間の研究論文（Research Article: RA）における言語的共起のパターンを明らかにした。Gray（2013）の研究は、研究論文における様々な状況的特徴（situational characteristics）を認識することで、学問分野内および学問分野を超えた学術的な言語に対する理解が深まることを示したものであると言える。Thompson *et al*（2017）は、学際

的な学術領域（interdisciplinary academic domain）の特徴を探るために、多次元分析を用い、環境学分野の 11,000 の学術論文からなるコーパスにおける言語変異を分析し、6 つの次元[18]を特定した。多次元分析の革新的な応用は、学際的な分野の複雑なディスコースを説明するのに特に効果的であると述べられている（Thompson *et al.* 2017）。Egbert（2015）は、6 つのジャンルにまたがる出版された学術的な文章に対して、多次元分析法で言語変異を探った結果、「Affective synthesis versus specialized information density」、「Definition and evaluation of new concepts」、「Author-centered stance」、「Reader-friendly narrative」、「Abstract observation and description」という 5 つの次元を特定できた。

また、アカデミックライティングにおける言語変異（linguistic variation within academic writing）に焦点を当てた研究も増えている。Jin（2018）では、工学分野（例、化学、工学）の英文研究論文（English research articles）の discussion 部分における言語的共起パターンと、インパクトの高い論文（high-impact articles）とインパクトの低い論文（low-impact articles）の discussion セクションを区別する言語変異について、多次元分析の視点で調査を行った。Parry（1998）では、博士課程学生のライティングにおける学問分野間の変異を考察し、科学・社会科学・人文科学という 3 つの学問分野において言語構造と引用の仕方を比較した結果、学問分野間の差異が確認された。Azher *et al*（2019）では、パキスタンの様々な大学の修士・博士課程の研究論文を分析データとして用い、パキスタンのアカデミックライティング（Pakistani academic writing）における分野間の言語変異を調査した。Friginal & Mustafa（2017）では、米国で出版された英語研究論文（research article）とイラク（Iraq）で出版されたイラク人著者による英語研究論文の要旨（abstracts）の言語特徴を調査している。その結果、米国人著者とイラク人著者が要旨をどのように構成するかについて、特に「情報のパッケージ化と共有の仕方」、「要旨における手続き的談話（procedural discourse）の表現」、「著者による直接性と論証の表現方法」について類似点と相違点があることが示唆された。

Omidian *et al*（2021）は、研究論文（research writing）における知識伝達（knowledge communication）の言語特徴を調査するために、10 分野（生物学、化学、歯学、機械工学、物理学、応用言語学、ビジネス、経営、政治、社会学）の学術誌に掲載された研究

[18] その 6 つの次元は (1) System-oriented vs. action-oriented、(2) Explicit vs. implicit argumentation、(3) Informality、 (4) Conceptual discourse、 (5) Text-focused vs. site-focused、 (6) Non-research world vs. research world である（Thompson *et al.* 2017）。

論文（約 450 万語）からなる学際的コーパスを分析対象とし、学問分野（disciplinarity）、テキスト内の変異（intra-textual variation）、L1 対 LX[19]の専門家の文章（L1 versus LX expert writing）という 3 つの要因に基づいて多次元分析を行った。分析の結果、研究論文における談話実践（discourse practices）の根底にある 3 つの基本的な次元が明らかになった。3 つの次元はそれぞれ、「Evaluative and elaborated discourse vs. Informational focus」、「Conceptual and abstract vs. concrete language」、「past-time procedural discourse」である。これらの機能的次元を定義する言語的共起パターンは、学問分野の種類に強く影響されうることが示されており、次元 1 （Evaluative and elaborated discourse vs. Informational focus）は L1 英語学者と LX 英語学者の執筆習慣を区別できることを明らかにした (Omidian *et al.* 2021) と述べられている。

(5) L2 ライティングの言語変異への応用

Biber の多次元分析モデルは言語変異の研究分野に大きく貢献しており、上記に述べたような英語の言語使用域だけではなく、外国語言語学習における言語変異の研究にも適用される。例えば、Xie（2020）は Biber（1988）によって特定された次元を用い、応用言語学の中国人学生が書いた修士論文の要旨の通時的変化を調査した。馬（2002）は 66 種の言語項目[20]を用い、中国人大学生とアメリカ人大学生が書いた英作文の言語特徴を比較分析した。その結果、中国人とアメリカ人大学生の英作文は、66 種の言語項目のうち 9 つの言語項目において有意差があり、中国人学生の英作文は二人称代名詞、談話助詞（discourse particles）、接続詞、形容詞が多く使われているのに対し、アメリカ人学生の英作文は比較的に長く、使役従属（causative subordinations）・動詞補語としての that 節（“that” clauses as verb complements）・形容詞補語としての that 節（“that” clauses as adjective complements）・関係詞節（relative clauses）などが多く見られることが分かった（馬 2002）。Goulart（2021）では、特定の L2 学生グループ（英国のブラジル人学部生）によって書かれたライティングが、英語を母語とするイギリス人大学生（L1）のライティングとどのように異なるのかを調べるために、L2 および L1 ライティングにおける言語変異の基礎となる次元（次元 1：expression of opinion versus compressed procedural information；次元 2：expression of possibility versus account of completed events；次元

[19] 「LX」は L2 や非ネイティブスピーカーを指している（Dewaele 2018）。

[20] 「文平均長」以外の 65 種の言語項目は Biber（1988: 222-245）によるものである。

3：informational density versus engaging presentations；次元 4：involved academic narrative versus elaborate descriptions；次元 5：stance towards the work of others）を抽出することができた。

Friginal & Weigle（2014）は Biber（1988, 1995, 2006）の多次元分析（MDA）手法 を、L2 ライティングにおける微細な言語変異に応用することを検討した。Friginal & Weigle（2014）は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）によって下記のような 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。

Dimension 1: Involved vs. Informational Focus

（次元 1：関与性 vs. 情報提示的言語生成）

Dimension 2: Addressee-Focused Description vs. Personal Narrative

（次元 2：宛先重視の記述 vs. 個人的物語調）

Dimension 3: Simplified vs. Elaborated Description

（次元 3：簡略的説明 vs. 詳細な説明）

Dimension 4: Personal Opinion vs. Impersonal Evaluation/Assessment

（次元 4：個人的な意見 vs. 非人称の評価/アセスメント）

以上の先行研究により、多次元分析は、ネイティブスピーカーと非ネイティブスピーカーの言語産出における言語変異を調べるのが可能であることが分かった。その他、非ネイティブスピーカーにおいても、言語習熟度レベルの異なる学習者の言語産出には違いがあることを示す研究も見られる。例えば、Kim & Nam（2019）は、Biber（1988）により得られた 5 つの次元を用い、異なる習熟度レベルの L2 ライティングにおける明確な言語的・機能的特徴のパターンを調査した結果、L2 のレベルが異なると、ライティングにおける言語的・機能的な特徴も異なることが分かった。具体的には、低習熟度グループのライティングは、よりネイティブスピーカーの話し言葉に近く、高習熟度グループのライティングは、よりネイティブスピーカーの書き言葉に近かったと結論づけている。

Pan（2018）は 3 つの習熟度レベル（修士論文、博士論文、研究論文）の L1 および L2 アカデミックライティング（academic writing）における言語変異を探るために、多次元

分析を行った。探索的因子分析により、L1 と L2 のアカデミックライティングにおける語彙および文法の違いを捉える 4 つの次元（次元 1: Attitudinal versus descriptive; 次元 2: Immediate style versus reported style; 次元 3: Academic involvement versus information density; 次元 4: Clausal style versus phrasal style）を特定して解釈ラベルを付けた。二元配置分散分析（Two-way ANOVA tests）を行った結果、4 つの次元のうち 3 つの次元（次元 1〜次元 3）で、L1 と L2 のアカデミック・ライティングに有意差があることを示した。L1 のアカデミック・ライティングと比較して、L2 のアカデミック・ライティングは一般的に、態度（attitudinal）、物語性（narrative）、学問的な関与（academically involved）が少ないことがわかったと述べられている（Pan 2018）。Pan（2018）の研究結果は、L1 と L2 のアカデミック・ライティングにおける異なるリソースの使用、および語彙と文法的特徴の選定が、最終的に彼らのライティングスタイルを形成していることを示唆している。

多次元分析の手法は、上記に述べた英語の使用域だけでなく、「英語以外の言語における言語変異の研究」（田畑 2005: 193）にも適用される。例えば、Asención-Delaney & Collentine（2011）は、大学 2 年生と 3 年生の学習者の L2 スペイン語ライティングコーパス（202,241 語）を対象に多次元分析を行った。その結果、2 年生と 3 年生の L2 スペイン語学習者のライティングにおける語彙・文法的特徴は、主に動詞または名詞的（verbal or nominal features）特徴を組み合わせた 5 つの次元に分類される（Asención-Delaney & Collentine 2011）と述べられている。また、近年では、多次元分析手法を使って、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も見られる。2.3 節では、中国語の言語変異に関する多次元分析の先行研究を整理する。

2.3 中国語の多次元分析

近年では、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（例、Zhang 2012; Zhu 2015; 劉 2019 等）、L2 中国語学習における言語変異の研究（例、徐 2021）などが見られる。例えば、Yanchun *et al*（2018）は Biber の多次元分析法を参考にして中国語のディベート（debate）と講演（speech）をレジスターの観点から比較分析した。Zhang（2012）は Lancaster Corpus of Mandarin Chinese（LCMC, McEnery & Xiao 2004）というコーパスを使用し、中国語の書き言葉に跨る 15 ジャンル（例、新聞報道・宗教・伝

記・公文書・一般小説・ミステリー小説等）間の言語変異を反映できる 3 つの次元を得ることに成功した。

Dimension 1：Literate

Dimension 2：Classical

Dimension 3：News commentary

Zhu (2015)は中国語の話し言葉と書き言葉に跨る 16 ジャンルを分析対象とし、探索的因子分析（Principal factor analysis）を用いて言語変異の反映された 5 つの次元を特定した。

Dimension 1：Interactive vs. Informational Discourse

Dimension 2：Literary vs. Non-literary Concern

Dimension 3：Colloquialized Expression with Subjective Emphases

Dimension 4：Situation-dependent Reference & Emotional Concern

Dimension 5：Persuasion and Argumentation vs. Non-persuasive and Non-argumentative Concern

徐（2021）は中国語母語話者の叙述文を比較対象とし、多次元分析で「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」・「行為の叙述」という 7 つの次元を捉え、日本人中国語学習者の叙述文における言語特徴を考察した。作文の例文を組み合わせながら分析した結果、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述べられている（徐 2021）。

2.4 本研究の位置付け

2.2 と 2.3 で整理した先行研究から分かるように、多次元分析の枠組みは、様々な言語使用域（例、書き言葉と話し言葉、ブログ、報道記事、学術文章、研究論文など）に応用することができ、特に、L1 または L2 ライティング内、あるいは L1 と L2 ライティング間の系統的な言語変異を捉えるためにうまく適用できることが示された。

L2 ライティング研究の分野では、多次元分析を用いて英語 L2 ライティングにおける言語変異を調査する研究は数多く見られるが、徐（2021）以外に、中国語 L2 ライティングにおける言語変異を多次元分析で調べる研究はほとんどない。本研究では、中国語母語話者が書いた中国語作文を参照対象とし、日本人中国語学習者によって書かれた作文における言語変異を多次元分析によって考察する。

また、Biber *et al*（2004）および Biber（2006）によれば、多次元分析には 2 つのタイプがある。一つ目のタイプは、既存の研究に特定された次元を使用し、新しいレジスターを特徴付けることである。例えば、Biber & Finegan（1989）は Biber（1988）で得られた 6 つの次元のうち、文語（literate）と口語（oral）の違いに関連する 3 つの次元（‘Informational versus Involved Production’、‘Elaborated versus Situation-Dependent Reference’、‘Abstract versus Nonabstract Style’）を用い、17 世紀から 20 世紀にかけての英語散文の 3 ジャンル（fiction、essays、letters）の史的推移を考察した。田畑（2002）は、Biber & Finegan（1989）に用いられた 3 つの次元を文体のベンチマークとして、上に述べた 17 世紀から 20 世紀にかけての英語散文と比較しながら、Dickens の文体の特徴を浮き彫りにした。Weigle & Friginal（2015）は Hardy & Römer（2013）の「Michigan Corpus of Upper-level Student Writing, MICUSP」というコーパスの分析から得られた 4 つの次元を用いて、英語母語話者と非英語母語話者が書いたテスト作文における言語特徴を比較した。しかし、話し言葉や書き言葉のレジスターのような一般的なコーパスから特定された次元は、特定の談話領域で実際に最も重要である特殊な機能の優先順位を反映していない可能性がある（Biber 2006: 179-181）と述べられている。

二つ目のタイプは新しい多次元分析と呼ばれ、調査対象の談話領域の言語的・機能的共起パターンを特定するために新たな多次元分析を行うことである。つまり、コーパスに頻繁に現れる言語特徴に基づく新しい因子分析を行い、新しい次元を生成するものである。例えば、前に述べた先行研究において、Asención-Delaney & Collentine（2011）、Egbert（2012）、Friginal & Weigle（2014）、Zhu（2015）などは、実際の調査対象に対して新たな因子分析を行い、新しい次元を特定してレジスターの変異を考察した。このアプローチは、多くの異なるテキスト・カテゴリーを含む新しい談話領域を分析する場合に適していると Biber *et al*（2004: 52-53）は述べている。

本研究の目的は、日本人中国語学習者が書いた中国語作文における言語特徴を、中国語母語話者の作文と比較し、言語変異の根本的な次元を特定することであるため、新し

い多次元分析を行う必要がある。従って、本研究では、前述に述べた先行研究における既存の次元をそのまま使わず、新たな多次元分析を行って、中国語母語話者と日本人中国語学習者の産出した書き言葉における言語変異を考察する。

第 3 章　言語項目の特定

本章では、先行研究における言語項目を整理して参考にした上で、本研究で選定した中国語言語項目の詳細、及び言語項目を集計するために開発したツールを紹介する。

3.1 言語項目

言語項目(Linguistic Features)[21]とは、多次元分析に用いる説明変数として分析に先立ち特定する一連の文法的・語彙的項目である。多次元分析の精度を高めるためには重要性の見込まれる言語項目（特定の伝達機能と関連づけられるもの）を可能な限り幅広く取り入れることが必要となる。例えば、Biber（1988）の多次元分析において用いられる言語項目は、Tense and aspect markers（時制と相の標識）、Place and time adverbials（場所と時間を表す副詞）、Pronouns and pro-verbs（代名詞と代動詞） など 16 種の文法カテゴリーに当てはまる計 67 言語項目（例えば、past tense、perfect aspect、present tense、place adverbials、time adverbials、first person pronouns、second person pronouns 等）である。言語項目の選定について、Biber（1988: 71-72）は下記のように述べている。

> Biber（1988: 71-72）
>
> Prior to any comparison of texts, a principled decision must be made concerning the linguistic features to be used. For the purposes of this study, previous research was surveyed to identify potentially important linguistic features - those that have been associated with particular communicative functions and therefore might be used to differing extents in different types of texts. No a priori commitment is made concerning the importance of an individual linguistic feature or the validity of a previous functional interpretation during the selection of features. Rather, the goal is to include the widest possible range of potentially important linguistic features.

[21] 「Linguistic Features」の日本語訳は「言語特徴」と「言語項目」がある。本論では「言語項目」という訳を使用する。

表 3.1 は、先行研究が特定された言語項目の一部を示している。

表 3.1　先行研究における言語項目の一部

文献	言語項目数	例
Bibe (1988)	67 linguistic features	1. past tense 2. perfect aspect 3. present tense 4. place adverbials 5. time adverbials … 65. independent clause coordination 66. synthetic negation 67. analytic negation: not
Biber (2006)	129 linguistic features	1. Type/token ratio 2. Mean word length 3. Distribution of word types by part-of-speech 4. Distribution of word types across registers 5. Distribution of word types across academic disciplines … 127. Quantity 128. Group / institution 129. Abstract / process
Zhang Z (2012)	60 linguistic features	1. noun 2. monosyllabic noun 3. verb used as noun 4. adjective used as noun 5. personal Name … 58. ‘(’: parenthesis 59. ‘;’: semi-column 60. ‘、’: pause used in enumerating items of a list (\)
Zhu (2015)	88 linguistic features	1. auxilliary *zhe* aspect article（“着”） 2. auxilliary *le* aspect article（“了”） 3. auxilliary *guo* aspect article（“过”） 4. place words 5. localizer … 86. quantifier verbs 87. quantifier temporal 88. onomatopoeia
徐 (2021)	111 linguistic features	1. Nouns: most commonly used 2. Nouns: moderately commonly used 3. Nouns: rarely used 4. Abstract nouns 5. Stance nouns …… 109. Lexical diversity 110. Average word length 111. Average sentence length

ある言語使用域(register)で言語項目の出現する頻度は、その言語使用域と関連する談話コンテキスト（discourse context）において、これらの項目がどの程度機能的に必要とされるかに依存する。なぜなら、言語項目は、関連する機能を持つために、テキストに共起する（Egbert & Biber 2018）からである。Biber（1988: 13）[22]が述べているように、「言語項目の強い共起パターンにより、その根底にある機能的次元が示される。言語項目はテキストの中でランダムに共起するわけではない。いくつかの項目が一貫して共起する場合、その使用を促す根底にある機能的な影響を探すのは妥当である」。言語項目の共起は、言語変異の一種とみなすことができる。

3.2 言語項目の特定

本研究では、多次元分析に関する先行研究（例、Biber 2006; Jack & Eric 2016; Zhang 2012; 劉 2019; 徐 2021 等）と、中国語の語彙や文法研究に関する専門書（例、黄・廖 2017; 馮 2000; 馮 2006 等）、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(付録Ⅱを参照されたい)を定めた。

92 種の言語項目における 43 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特徴（例えば、抽象名詞、第二人称代名詞、語彙多様性[23]等）である。中国語には英語のような節（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・廖（2017: 129-139）で考察されている 10 種類の複文（言語項目の ID：79-88）は分析に含まれている。また、アカデミック的中国語と非アカデミック的中国語に関する研究(Zhu 2015; Zhang 2012; 劉 2019 など）を参考にして、 35 種の言語項目（例：着/zhe、了 /le、过/guo、的 /de、地 /de、得 /de、韻律単音節語、韻律双音節語等）が言語項目の構成要素として追加されている。そのほかに、本研究の研究対象は日本人中国語学習者の中国語ライティングであるので、「HSK 語彙リスト（合計 5000 語）」に基づき、学習者の語彙使用能力を反映できる 4 種の言語項目、すなわち、

[22] “[S]trong cooccurrence patterns of linguistic features mark underlying functional dimensions. Features do not randomly co-occur in texts. If certain features consistently co-occur, then it is reasonable to look for an underlying functional influence that encourages their use” (Biber 1988: 13).

[23] Biber（2006）において、「抽象名詞」、「第二人称代名詞」と「語彙多様性」に対応する英語の言語項目名はそれぞれ、 “Common nouns: abstract”、“Pronouns: 2nd person”、“Type/token ratio”である。

低難度語彙（HSK1-2 級の語彙）・中難度語彙（HSK3-4 級の語彙）・高難度語彙（HSK5-6 級の語彙）・非 HSK 語彙（HSK1-6 級に収録されていない語彙）[24]も本研究の言語項目のリストに含めている。

付録 II には、本研究の分析に使用される 92 種の言語項目とその ID を示している。これら 92 種の中国語言語項目の詳細は下記の通りである。

まず、付録 II に示した ID 番号 1-3、11-13、23-25、29-31 を表す言語項目（高頻度名詞、中頻度名詞、低頻度名詞、高頻度動詞、中頻度動詞、低頻度動詞、高頻度形容詞、中頻度形容詞、低頻度形容詞、高頻度副詞、中頻度副詞、低頻度副詞）は、Biber（2006: 182, 184-185）による 12 種の言語項目である。Biber（2006）におけるこれら 12 種言語項目の対応する英語名は以下の通りである。

ID1: 高頻度名詞 → high frequency nouns（Biber 2006:182）
ID2: 中頻度名詞 → nouns: moderately common（Biber 2006:184）
ID3: 低頻度名詞 → nouns: rare（Biber 2006:184）
ID11: 高頻度動詞 → high frequency verbs（Biber 2006:182）
ID12: 中頻度動詞 → verbs: moderately common（Biber 2006:184）
ID13: 低頻度動詞 → verbs: rare（Biber 2006:184）
ID23: 高頻度形容詞 → high frequency adjectives（Biber 2006:182）
ID24: 高頻度形容詞 → adjectives: moderately common
ID25: 高頻度形容詞 → adjectives: rare（Biber 2006:184）
ID29: 高頻度副詞 → adverbs: common（Biber 2006:184）
ID30: 高頻度副詞 → adverbs: moderately common（Biber 2006:184）
ID31: 高頻度副詞 → adverbs: rare（Biber 2006:184）

[24] これらの 4 つの言語項目は HSK 語彙リスト（計 5000 語）に基づいて特定されたものである。HSK は （中国語名: 漢語水平考試）は中国の教育部が認定する国際的な中国語の語学検定試験である。HSK の公式サイトには、1 級（最低級）~ 6 級（最高級）で使用される語彙（計 5000 語）のリストが《新汉语水平考试(HSK)词汇(2012 年修订版)》の名称で配布されている。URL: https://www.chinesetest.cn/godownload.do〔最終アクセス日:2022 年 5 月 20 日〕

上記の 12 種の言語項目に属する中国語の語彙は、主に中国語“cn-corpus”の“现代汉语语料库分词类词频表”[25]の語彙リストを参考にし、名詞・動詞・形容詞・副詞の高頻度・中頻度・低頻度語彙リストとして特定されたものである。図 3.1 は“cn-corpus”ホームページの画面を示している。

図 3.1 “语料库在线 www.cncorpus.org”のホームページ

“现代汉语语料库分词类词频表”（以下は「品詞頻度表」とする）は、2000 万字のコーパスにおいて 50 回以上出現した品詞別の単語（計 16,254 異なり語数）を集計した語彙表である。「品詞頻度表」において、各種類の品詞は出現回数（粗頻度）の高い順に並べており、50 回以上出現する名詞、動詞、形容詞、副詞の異なり語数はそれぞれ 6,508、5,219、1,463、569 である。図 3.2 は「品詞頻度表」リストの一部、図 3.3 は「品詞頻度表」における名詞リストの一部を示している。

[25] “现代汉语语料库分词类词频表”（現代中国語コーパスの品詞頻度表）：http://corpus.zhonghuayuwen.org/Resources.aspx（2022.03.19 最終アクセス）

図 3.2 と図 3.3 に示したエクセルの 1～4 行目はそれぞれ、「現代中国語コーパス品詞頻度表」、「コーパスサイズ：2000 万字」、「50 回以上出現する単語のみを収録している」、「www.cncorpus.org よりダウンロードしたものである」という意味で、7 行目の A～G 列はそれぞれ「id」、「単語」、「タグ」、「品詞名」、「出現回数（粗頻度に相当する）」、「頻度（相対頻度に相当する）」、「累積頻度」を意味する。

	A	B	C	D	E	F	G
1	#现代汉语语料库词语分词类频率表						
2	#语料规模:2000万字						
3	#只列入出现次数大于50次的词						
4	#下载自www.cncorpus.org语料库在线网站						
5							
6							
7	id	词语	词类标记	词类名称	出现次数	频率	累积频率
8	1	的	u	助词	744863	7.7946	7.7946
9	2	了	u	助词	129617	1.3564	9.151
10	3	是	vl	联系动词	118382	1.2388	10.3898
11	4	在	p	介词	115425	1.2079	11.5977
12	5	和	c	连词	79845	0.8355	12.4332
13	6	一	m	数词	78590	0.8224	13.2556
14	7	这	r	代词	65143	0.6817	13.9373
15	8	有	v	动词	53522	0.5601	14.4974
16	9	他	r	代词	52908	0.5537	15.0511
17	10	我	r	代词	52724	0.5517	15.6028
18	11	也	d	副词	47034	0.4922	16.095
19	12	不	d	副词	46950	0.4913	16.5863
20	13	就	d	副词	44145	0.462	17.0483
21	14	着	u	助词	40596	0.4248	17.4731
22	15	中	nd	方位名词	40105	0.4197	17.8928
23	16	地	u	助词	39440	0.4127	18.3055
24	17	说	v	动词	35047	0.3667	18.6722
25	18	上	nd	方位名词	34850	0.3647	19.0369
26	19	都	d	副词	34261	0.3585	19.3954
27	20	人	n	名词	33915	0.3549	19.7503
28	21	个	q	量词	31445	0.3291	20.0794
29	22	把	p	介词	28786	0.3012	20.3806
30	23	你	r	代词	28769	0.3011	20.6817
31	24	种	q	量词	28587	0.2991	20.9808
32	25	对	p	介词	28397	0.2972	21.5752
33	26	而	c	连词	28397	0.2972	21.278
34	27	要	vu	能愿动词	27324	0.2859	21.8611
35	28	我们	r	代词	26821	0.2807	22.1418

図 3.2　“现代汉语语料库分词类词频表”の一部（2022.03.19 最終アクセス）

	A	B	C	D	E	F	G
1	#现代汉语语料库词语分词类频率表						
2	#语料规模:2000万字						
3	#只列入出现次数大于50次的词						
4	#下载自www.cncorpus.org语料库在线网站						
5							
6							
7	id	词语	词类标记	词类名称	出现次数	频率	累积频率
27	20	人	n	名词	33915	0.3549	19.7503
80	73	社会	n	名词	11461	0.1199	29.9738
84	77	问题	n	名词	10899	0.1141	30.4416
90	83	国家	n	名词	10138	0.1061	31.0897
92	85	工作	n	名词	9655	0.101	31.295
104	97	生活	n	名词	8694	0.091	32.4398
105	98	经济	n	名词	8680	0.0908	32.5306
119	112	关系	n	名词	7715	0.0807	33.7232
120	113	人们	n	名词	7702	0.0806	33.8038
123	116	作用	n	名词	7548	0.079	34.043
133	126	我国	n	名词	6948	0.0727	34.8089
135	128	情况	n	名词	6922	0.0724	34.9538

図 3.3　名詞リストの一部（2022.03.19 最終アクセス）

本研究では、上記に紹介した「品詞頻度表」の名詞・動詞・形容詞・副詞リストにおける粗頻度の高い順の「0～25%、25%～75%、75%～100%」部分の単語を「高頻度、中頻度、低頻度」の名詞・動詞・形容詞・副詞として特定した。これら 12 種の言語項目の ID は連続していない（ID： 1-3、11-13、23-25、29-31）が、同じような方法で選択されたので、下記にまとめて紹介する（その後、他の言語項目は付録 II で示した ID の順に紹介する）。これら 12 種の言語項目に属する単語の上位 30 語は以下の通りである。

言語項目 ID： 1-3、11-13、23-25、29-31

ID1_高頻度名詞：人、社会、问题、国家、工作、生活、经济、关系、人们、作用、我国、情况、社会主义、人民、方面、时候、条件、思想、水、过程、科学、方法、技术、时间、事、世界、企业、物质、同志、话

ID2_中頻度名詞：胡子、年级、微生物、军官、地主阶级、公园、我军、图案、老汉、潮、奖金、各族、书籍、少女、模样、葡萄、圣经、建筑物、教员、儿女、变异、肩膀、证据、面孔、伯伯、大自然、厨房、品德、套、泡

ID3_低頻度名詞：歌手、下一代、传统观念、系统工程、创汇、原始人、结核、切片、收获、厂矿、师团、岩溶、呼吸道、碳酸钠、高技术、清军、岁数、萨

满教、读数、富农、决赛、西红柿、写字台、自来水、言行、天子、这般、布景、官府、疟疾

ID11_高頻度動詞：有、说、到、发展、使、去、没有、看、进行、如、生产、用、想、研究、知道、做、成、走、需要、发生、叫、吃、具有、形成、影响、认为、产生、出现、作、起

ID12_中頻度動詞：送来、被迫、停留、管辖、披、修理、没人、赴、协助、遇、乐、会、背着、现、造型、应、公布、刻、沉积、培育、置、缩、垂、征服、萌发、阐述、杂交、忍不住、告、用力

ID13_低頻度動詞：交易、高举、休养、正是、毕、取下、潜入、领悟、收费、夹杂、操心、活跃、构、掘、插嘴、试点、评比、野生、递增、算得、擅长、默、不在乎、化装、受理、谅解、荣获、背离、只管、新婚

ID23_高頻度形容詞：大、小、多、好、新、主要、不同、重要、一般、许多、高、一样、长、基本、直接、具体、正确、老、全、一定、整个、少、快、低、大量、自然、容易、严重、自由、所谓

ID24_中頻度形容詞：舒服、最低、得意、随便、专、片面、英勇、平常、暖、可惜、不远、单一、迟、富裕、熟练、明亮、寂寞、整齐、酸、足、不满、有用、剧烈、古典、饿、可见、理性、欢乐、勉强、有益

ID25_低頻度形容詞：流畅、任意、最优、无用、疲乏、惊恐、无能、轻盈、凉爽、空洞、无耻、基、可取、深邃、阴沉、飞速、曲、光、危急、牢、清澈、脆弱、两样、勇、明朗、良性、最终、不均、质朴、幸运

ID29_高頻度副詞：也、不、就、都、又、还、很、就是、更、已、却、再、最、才、已经、必须、便、只、较、将、没有、没、即、真、一定、还是、特别、先、曾、正

ID30_中頻度副詞：实在、不必、皆、不得不、快、好、一旦、并、至今、稍、早就、更为、至少、不过、由此、并非、基本上、反复、不能不、尚、果然、极其、紧紧、凡是、挺、纷纷、再也、反而、从此、颇

ID31_低頻度副詞：重点、一概、一举、甚至于、亲手、彼此、分外、再度、分明、按时、莫非、一味、岂不、连声、立时、宁可、亲身、果真、否则、差不多、实、最好、反倒、好容易、岂不是、大抵、骤然、白白、恰

言語項目 ID： 4-10

「抽象名詞、具象名詞、心理名詞、指人名詞、集合名詞」は Biber（2006: 184-185）[26]によるものである。「普通名詞の複数形」は王（2017）[27]による言語項目であり、「名詞化機能詞」は Zhu（2015: 74）[28]によるものである。下記では、抽象名詞・具象名詞・心理名詞・指人名詞・集合名詞に属する語彙の一部（30 語）を示している。

ID4_抽象名詞：哀荣、爱好、暗流、奥秘、把柄、把势、把戏、罢免权、霸权、白案、白班、白描、白眼、百年大计、板眼、版权、半边、梆子、宝号、保护伞、保守主义、暴力、杯水车薪、背水阵、被选举权、奔头、闭门羹、奥秘、魅力、底线

ID5_具象名詞：白点、百货、宝物、本身、舶来品、参照物、残毒、残骸、残余、残渣、草刺、柴米、陈列品、陈设、储备、传家之宝、搭头、单骑、单字、抵押物、附属物、附着物、公馆、公物、贡品、古物、景物、婴孩、酸奶、取景框

ID6_心理名詞：爱心、冰霜、惧高症、恐怖症、恐惧症、时代感、童真、心境、心事、抑郁症、躁狂症、爱情、爱憎、悲欢、博爱、恻隐之心、成就感、痴情、愁绪、愁云、春心、春意、单相思、妒忌心、父爱、负罪感、孤独感、光荣感、积怨、激情

ID7_指人名詞：老爸、老妈、阿斗、阿飞、阿公、阿訇、阿婆、阿 Q、阿姨、矮子、爱将、爱侣、爱人、暗娼、暗探、八路、把兄弟、爸爸、霸王、霸主、白

[26] Biber (2006: 184-185): ‘common nouns: abstract’, ‘common nouns: concrete’, ‘common nouns: mental’, ‘common nouns: human’, ‘common nouns: group’.

[27] 王永娜（2017）.《汉语书面正式语体语法的泛时空化特征研究》. 海外华文教育动态, (2), 109-109.

[28] Zhu (2015: 74): ‘verbs functioning as noun’, ‘adjectives functioning as noun’.

眼狼、白眼儿狼、败子、败家子、班主、半彪子、半熟脸、帮佣、帮主、包青天

ID8_集合名詞：编委会、党代会、藩国、藩属、妇代会、妇救会、妇联、敢死队、干部科、港务局、高校、歌剧院、歌舞团、革命军、革委会、革新派、各族、工厂、工会、工联、工农、工农兵、工农联盟、工人阶级、工商联、工学院、工业部、工作组、公安部、公安局

ID9_普通名詞の複数形：普通名詞＋“们”

ID10_名詞化機能詞：巧妙/a 的/ude1 <u>融合</u>/vn、有/uyou 意义/n 的/ude1 <u>研究</u>/vn、先生/n 的/ude1 <u>稳重</u>/an、给予/vx <u>方便</u>/an

名詞化機能詞（名詞化とも呼ばれる：nominalization）とは、動詞や形容詞として表現した方が適切なものを、名詞又は名詞句の形で表現することである。名詞化は、長い説明を名詞や複雑な名詞句に凝縮することを可能にすると述べられている（Schleppegrell 2001）。Chafe（1985: 108）[29]では、名詞化は「*tend, prefer, speak, refer, use* などの動詞や *abstract* などの形容詞が名詞句になり、他の動詞の引数や前置詞の対象になる」プロセスであると定義されている。また、上記の「ID10_名詞化機能詞」に示した例のように、NLPIR タグセットにおいて、動詞と形容詞の名詞的用法をそれぞれ「vn」と「an」タグで付与するため、「名詞化機能詞」の出現頻度を「vn」と「an」タグで集計することができる。

言語項目 ID： 14-22

「動作動詞、心理動詞、肯定を表す動詞、伝達動詞、推測を表す動詞」は Biber（2006:184）[30]によるものである。下記に示したのは、これら 5 つの言語項目に属する語彙の一部である。「方向動詞、副動詞、軽動詞、系動詞‘是’」は

[29] Chafe (1985: 108): “[B]y which verbs like *tend*, *prefer*, *speak*, *refer*, and *use* or adjectives like *abstract* become noun phrases that can then be the arguments of other verbs or the objects of prepositions.’’

[30] Biber (2006: 184): ‘common verbs: activity’, ‘common verbs: mental’, ‘that-clauses: controlled by certainty verbs’, ‘common verb: communication’, ‘that clauses: controlled by likelihood verbs’

Zhu（2014: 74-76）[31]による言語項目であり、NLPIR タグセットにおいて、それらの言語項目に対応するタグはそれぞれ、‘/vf’、‘/vd’、‘/vx’、‘/vshi’である。

ID14_動作動詞：哀叹、唉声叹气、安装、按摩、昂首阔步、拔草、拔除、拔掉、拔锚、把守、摆放、摆弄、摆手、伴唱、伴舞、伴奏、绑架、包扎、包装、剥、保管、保卫、抱、刨、背书、背诵、播种、唱歌、拔掉、拆迁

ID15_心理動詞：哀怜、爱怜、爱慕、爱屋及乌、爱惜、按捺、暗喜、暗笑、黯然神伤、懊悔、懊恼、懊丧、百感交集、百思不解、百无聊赖、半惊半喜、半信半疑、包涵、抱恨、悲天悯人、悲痛欲绝、悲喜交集、鄙视、鄙夷、闭门思过、变心、忏悔、揣摩、猜想、猜疑

ID16_肯定を表す動詞：断定、推论、证实、证明、说明、确定、查明、发现、发觉、了解、理解、支持、得知、看到、听到、观察、明白、知道、承认、认识、认可、记起、记得、领会、表明、懂得

ID17_伝達動詞：哀告、哀求、安慰、帮腔、禀报、禀告、禀明、嘲讽、嘲弄、嘲笑、称呼、嗤笑、耻笑、斥骂、斥责、传呼、传唤、传讯、答复、点拨、点将、刁难、对唱、对答、敦促、敦请、敦劝、反诘、反诉、反问

ID18_推測を表す動詞：猜、估算、猜测、猜度、猜想、臆想、假定、认为、相信、以为、怀疑、疑惑、推断、推测、想象、思索、看来、觉得

ID19_方向動詞：来/vf、去/vf、起来/vf、下去/vf、出来/vf、出去/vf

ID20_副動詞：<u>持续</u>/vd 增长/v、<u>带头</u>/vd 违反/v 纪律/n

ID21_軽動詞：<u>进行</u>/vx 研究/vn、<u>给予</u>/vx 宝贵/a 意见/n、<u>加以</u>/vx 改进/v

ID22_系動詞‘是’：他/rr <u>是</u>/vshi 学生/n

言語項目 ID： 26-28

「性質形容詞」は黄・廖（2017 下冊: 12）の“性质形容词”によるもので、「状態形容詞」と「区別詞」は Zhu（2014: 75）[32]による言語項目である。それら 3 つの

[31] Zhu (2014: 74-76): ‘directional verbs’, ‘verbs used as adverb’, ‘light verb’, ‘verb shi (be) as main verb’.

言語項目のタグはそれぞれ、‘/a’、‘/z’、 ‘/b’である。各言語項目に属する語彙の一部は下記に示している。

ID26_性質形容詞：好、坏、伟大、勇敢、优秀、聪明、大方、大、小、高、低、长、短、肥、瘦

ID27_状態形容詞：雪白、笔直、墨绿、火热、血红、绿油油、水灵灵、黑不溜秋、冰冷、沉甸甸、孤零零

ID28_区別詞：副/b 校长/n、野生/b 动物/n

言語項目 ID： 32-37

副詞は動詞や形容詞を修飾し、程度・範囲・時間などの意味を示す。「必然性を表す副詞、可能性を表す副詞、態度を表す副詞、否定を表す副詞、時間を表す副詞、程度を表す副詞」は、黄・廖（2017:18-19）における副詞の分類（表 3.2）、崔（2002）と柳（2007）の博士論文[33]及び宋(2008)の副詞に関する研究に基づいて整理された言語項目である。

[32] Zhu (2014: 75): ‘adjectives functioning as the attributes of the nouns and predicates’, ‘adjectives functioning as the modifiers of the nouns’.

[33] 崔诚恩(2002). 现代汉语情态副词研究(Doctoral dissertation, 中国社会科学院).
柳燕妮(2007). 副词的分类与释义分析(Doctoral dissertation, 西北师范大学).
宋晓娟(2008). 时间副词用于肯定与否定的对称与不对称现象的研究（Master's thesis, 陕西师范大学).

表 3.2　中国語副詞の分類[34]

副词限制、修饰动词、形容词性词语，表示程度、范围、时间等意义。（訳：副詞は動詞や形容詞を制限・修飾し、程度・範囲・時間などの意味を表すものである。）

(1)表示程度(程度を表す)：很、最、极、挺、太、非常、十分、极其、格外、分外、更、更加、越、越发、有点儿、稍、稍微、略微、几乎、过于、尤其、倍加、不大、比较、差不多

(2)表示范围(範囲を表す)：都、大都、均、总、共、总共、统共、只、仅仅、单、净、光、一齐、一概、一律、单单、不单、就、全、通通、皆

(3)表示时间、频率(時間・頻度を表す)：已、已经、曾、曾经、刚、才、刚刚、正、在、正在、将、将要、就、就要、马上、立刻、顿时、终于、常、常常、时常、时时、往往、渐渐、早晚、一向、向来、总是、始终、永、赶紧、仍然、还是、屡次、依然、重新、还、再、再三、偶尔

(4)表示处所(場所を表す)：四处、随处。

(5)表示肯定、否定(肯定・否定を表す)：必、必须、必定、准、的确、不、不必、没有、没、未、未必、别、莫、勿、是、是否、不必、不用(甭)、不曾、未曾、未尝、无须、无(毋)庸。

(6)表示方式、情态(様態・情態を表す)：大肆、肆意、特意、猛然、忽然、公然、连忙、赶紧、悄悄、暗暗、大力、稳步、阔步、单独。

(7)表示语气(語気を表す)：难道、岂、究竟、偏偏、索性、简直、就可、也许、难怪、大约、幸而、幸亏、反倒、反正、果然、居然、竟然、必然、必须、确实、何尝、何必、明明、恰恰、未免、只好、不妨、根本。

(8)表示关联(関連性を表す)：便、也、又、却、再、就。

注：黄・廖（2017 下冊: 18-19）における副詞の分類に基づいて作成

ID 32-37 の各言語項目に属する副詞の語彙は下記の通りである。

ID 32_必然性を表す副詞： 毋庸置疑、显而易见、无疑、毫无疑问、必、必须、必定、准、不、没有、没、未、未必、别、莫、勿、是、是否、不必、不用、甭、准、不曾、未曾、无须、无庸、的确、确实、实在、当然、显然、诚然、根本、一定、必然、想必、势必、总归、准保

[34] 括弧内にある内容は日本語の訳である。

ID 33_可能性を表す副詞： 也许、或许、兴许、可能、恐怕、好像、仿佛、宛然、不大、比较、差不多、大概、大约、或多或少、几乎、稍微、些许、有些、似乎、略微、稍、大致

ID 34_態度を表す副詞： 难道、岂、究竟、偏偏、索性、简直、也许、难怪、大约、幸而、幸亏、反倒、反正、果然、居然、竟然、必然、必须、确实、何尝、何必、明明、恰恰、未免、只好、不妨、根本、大肆、肆意、特意、猛然、忽然、公然、连忙、赶紧、悄悄、暗暗、大力、稳步、阔步、单独、无妨、未必、幸好、好在、多亏、何不、莫非

ID 35_否定を表す副詞： 不、不必、没有、没、未、未必、别、莫、勿、是否、不用、甭、不曾、未曾、未尝、无须、毋庸、无庸

ID 36_時間を表す副詞： 本来、比比、比年、比岁、不断、不日、不时、曾经、长年、长夜、常常、常川、常年、彻夜、趁早、成年、成日、成天、成宿、成夜、迟迟、迟早、重新、垂垂、从此、从来、从头、从小、从新、猝尔、猝然、打头、当即、当下、当时、登时、迭次、动辄、陡然、顿然、顿时、俄而、俄尔、反复、方才、方将、方始、纷纷、改日、改天、赶紧、赶快、赶忙、赶巧、赶早、刚刚、刚好、刚巧、根本、跟着、姑且、古来、惯常、过天、还是、忽地、忽而、忽然、回头、霍地、霍然、及时、及早、即将、即刻、即时、几曾、见天、间或、渐次、渐渐、将近、将要、届期、届时、尽管、尽快、尽早、尽自、经常、竟日、久久、久己、逮然、可巧、溢然、快要、老是、累次、累累、历来、立地、立即、立刻、立马、立时、聊且、临了、陆续、屡次、屡屡、麻利、马上、每常、每每、猛可、猛然、偶尔、偶或、偶然、频频、恰好、恰恰、恰巧、权且、仍旧、仍然、日夕、生来、时常、时而、时刻、时时、实时、始终、适才、首先、率先、素来、素昔、随即、随时、往往、未几、无何、无日、兀自、先后、向来、新近、行将、旋即、迅即、奄忽、奄然、眼见、眼看、要紧、业经、业己、一划、一旦、一动、一度、一径、一举、一连、一路、一时、一下、一向、一再、一朝、一直、依旧、依然、己而、已经、已然、应时、永世、永续、永远、犹自、有时、于今、预先、原本、原来、再次、再度、再三、在先、暂且、早日、早晚、早己、朝夕、正好、正巧、正在、至今、终年、终日、终岁、终天、终于、骤然、逐渐、自古、自来、抽冷子、到了儿、到头来、

动不动、赶明儿、冷不丁、冷不防、猛不丁、猛不防、猛孤丁、起小儿、时不时、压根儿、一会儿、一小儿、早早儿、本、便、必将、不大、才、曾、常、初、重、从、匆匆、凑巧、到底、都、徒然、方、刚、还、忽、渐、将、就、就要、快、老、屡、蓦地、蓦然、频、顷刻、且、仍、日见、日渐、日益、尚、时、素、随后、先、现、要、已、又、再、暂、暂时、在、早、乍、照常、照样、照旧、正、直、终、逐步、逐年、逐日、总、骤、总是、早就、业已

ID 37_程度を表す副詞： 很、最、太、极、更、顶、挺、较、越、愈、颇、大、好、多、超、超多、十分、多么、有点儿、差不多、几乎、稍、稍微、略微、非常、分外、格外、过于、极度、极其、相当、比较、更加、越发、尤其、甚、老、特、大为、何等、极端、颇为、万分、最为、越加、愈加、过分、特别、异常、较为、好不、何其、更为、尤为、甚为、怪、极为、不大、愈发、过、蛮、绝、尤、满、至、无比、不胜、愈益、愈为、透、顶顶、绝对、好生、倍加、忒

言語項目 ID： 38-46

代名詞に関する言語項目は 9 種類であり、一人称・二人称・三人称に関する人称代名詞、及び指示代名詞・不定代名詞・疑問代名詞などを含む。これらの 9 種の言語項目は Biber（2006）と Zhang（2012）によるものである。

Biber（2006: 184-185）に示した言語項目において、“pronouns: 1st person”、“pronouns:2nd person”、“pronouns: 3rd person”、“pronouns: it”、“pronouns: demonstrative”、“pronouns: indefinite”、“WH questions”[35]などを設けている。しかし、Biber（2006）は “pronouns: 1st person” の単数と複数を区別していない。中国語の多次元分析の研究において、Zhang（2012）は、レジスターにおける**“我”**（訳：私）と**“我们”**（訳：私たち）の分布パターンに違いがあり、**“我”**と**“我们”** の間には負の相関（相関係数：-0.195）があることを報告している。例えば、**“我”**（一人称単数代名詞）はすべての大衆言語レジスター（popular registers）でより頻繁に使われているようだが、**“我们”**

[35] “WH questions” は “what”、 “when”、“where”、“who”、“whom”、“which”、“whose”、“why”、“how”などの WH 単語で始まる疑問文である。中国語文脈における疑問代名詞（中国語：疑问代词）は英語疑問文の WH 単語に対応すると考えられる。

（一人称複数代名詞）は主にニュースレジスター（news registers）で多く使われているようである（Zhang 2012）。また、L2 中国語ライティングの研究分野において、馬（2017）と徐（2021）は、中国語母語話者より中国語学習者によって書かれた中国語作文の方が、一人称代名詞**“我”**と**“我们”**の出現頻度が高いことを述べている。そのため、「一人称代名詞」を「一人称単数代名詞“我”」、「一人称複数代名詞“我们”」、「その他の一人称代名詞」に細分化する必要があると考えられる。

ID 38_一人称単数代名詞“我”： 我

ID 39_一人称複数代名詞“我们”： 我们

ID 40_他の一人称代名詞[36]： 咱、咱们、本人、吾、余、予、愚、鄙人、敝人、不才、臣、私、身、本席、俺、俺们、俄、俄们、我伲、阿拉、阮、涯、偶、禾、昂、洒家、朕、予一人、孤、孤家、不谷、寡人、哀家、本宫、寡小君、小童、梓童、本座、本官、本将、下官、末将、在下、仆、小人、小弟、愚兄、小妹、愚姐、小生、晚生、后学、老夫、老朽、老身、老奴、微臣、臣妾、妾身、贱妾、小女、奴家、奴婢、贫僧、贫尼、贫道、草民、小的、某甲、某乙、自个儿、侬、侬家、奴、阿奴、儿、儿家、尔公、老子、人家、老娘、乃公、卑职、小可、不佞

ID 41_二人称代名詞： 你、您、你们

ID 42_三人称代名詞： 她、他、她们、他们

ID 43_代名詞“它(们)”： 它[37]、它们

ID 44_指示代名詞： 这、那、这会儿、那会儿、这儿、那儿、这里、那里、这样、那样、这么样、那么样、这么、那么、每、各、某、本、另、该、别的、其他、其余、其、彼、此、这些、那些、这边、那边

ID 45_不定代名詞： 有人、什么、哪儿、多会儿、怎么、某事物、某事、某物、某人、某、没有一个人、无论何事、任何事物、任何人、任何事、任何

---

[36] 参考：http://www.360doc.com/content/10/0310/23/710518_18297696.shtml〔最終アクセス日：2022 年 9 月 10 日〕

[37] 中国語の“它”は人間以外の事物を指す三人称代名詞である。

ID 46_疑問代名詞：谁、什么、哪、哪儿、哪里、怎么、怎样、怎么样、为什么

言語項目 **ID： 47-51**

「数詞、量詞、感嘆詞、擬音語」は Zhu（2014: 76）[38]によるものであり、タグはそれぞれ、‘/m’、‘/q’、‘/e’、‘/o’である。「方位場所詞」は Biber（2006: 184）の‘adverbials: place’、及び NLPIR のタグセットにおける‘/f’[39]と‘/s’[40]（中国語：处所词）によって設けられた言語項目である。各言語項目に属する語彙の例を下記に示している。

ID 47_数詞： 一、20、千、一些、一个

ID 48_量詞： 两点、一分钟、十元、六只、来一趟

ID 49_方位場所詞： 上/f、下/f、窗外/s、心中/s

ID 50_感嘆詞： 唉/e、嗨/e、哎呀/e、哟/e

ID 51_擬音語： 嘭/o、轰隆轰隆/o、叽叽喳喳/o、咩/o、喵/o

言語項目 **ID： 52-58**

「列挙助詞“等/等等”」と「比況助詞」は Zhu（2014: 75）による言語項目であり、「構造助詞“的”、構造助詞“地”、構造助詞“得”、時態助詞“着”、時態助詞“了”、時態助詞“过”」は Zhu（2014）と Zhang（2012）による言語項目である。Zhang（2012）[41]によれば、「中国語には 3 つの同音異義語構造マーカー（structural markers）があり、それぞれ名詞修飾、副詞修飾、動詞補語を示す de と発音する」、「中国語は時制を表さないが、完了、進行/持続、経験を表すマーカーがある（それぞれ*-le*、*-zhe*、*-guo*）」と述べられている。

---

[38] Zhu (2014: 76): ‘numerals, ‘quantifiers’, ‘exclamation’, ‘onomatopoeia’.

[39] ‘/f’は方位を表す語彙（中国語：方位词）のタグである。

[40] ‘/s’は場所を表す語彙（中国語：处所词）のタグである。

[41] Zhang(2012): “There are three homophonous structural markers in Chinese 的, 地, and 得, all pronounced ***de***, for marking nominal modification, adverbial modification and verbal complementation respectively.” “While Chinese does not mark tense, it does have markers for perfective, progressive/durative and experiential aspects (***-le***, ***-zhe*** and ***-guo***, respectively).”

構造助詞において、構造助詞“的”は所属や修飾関係を表す機能を持っている。構造助詞“地”は連用修飾語を明示することができる[42]。構造助詞“得”は結果／程度／可能などに関わる補語を明示することに用いられる。

時態助詞とはアスペクトマーカー（Aspect marker）とも呼ばれ、動詞の直後に置いて動作の「“着”：継続、～している」・「“了”：実現・完了、～した」・「“过”：経験、～したことがある」を表現する助詞である[43]。

Zhang（2012）と Zhu（2014: 74-75）では、これら 8 種言語項目に対応する英語名は下記の通りである。

ID52_構造助詞“的”：Nominal ***de*** (Zhang 2012); auxiliary ***de*** used after the noun for the possessive case of noun (Zhu 2014: 75)

例文：我<u>的</u>书（私の本）、幸福<u>的</u>生活（幸せな生活）

ID53_構造助詞“地”：Verbal modification ***de*** (Zhang 2012); auxiliary ***de*** used after the adverb to link the adverb and verbs (Zhu 2014: 75)

例文：好好（<u>地</u>）工作（ちゃんと仕事をする）、
慢慢（<u>地</u>）说（ゆっくり話す）

ID54_構造助詞“得”：Verbal complement ***de*** (Zhang 2012); auxiliary ***de*** used between the adverbs and the verb phrases expressing the results of the adverbs (Zhu 2014: 75)

例文：做<u>得</u>到（やり遂げる）、起<u>得</u>很早（起きるのがとても早い）、
听<u>得</u>懂（聞いて理解できる）

ID56_時態助詞“着”：progressive/durative ***zhe*** (Zhang 2012);
auxiliary ***zhe*** aspect article (Zhu 2014: 74)

例文：他在椅子上坐<u>着</u>。（彼は椅子に座っている）、
他穿<u>着</u>西服。（彼はスーツを着ている）

ID57_時態助詞“了”：perfective ***le*** (Zhang 2012); auxiliary ***le*** aspect article (Zhu 2014: 74)

例文：她买<u>了</u>书。（彼女は本を買った。）、

---

[42] https://www.chinesemaster.net/grammar/%E6%A7%8B%E9%80%A0%E5%8A%A9%E8%A9%9E〔最終アクセス日：2022 年 10 月 5 日〕

[43] https://china.younosuke.com/03_050.html〔最終アクセス日：2022 年 10 月 5 日〕

今天他吃了早饭。（今日彼は朝食を食べた。）

ID58_時態助詞“过”：experiential aspects ***guo*** (Zhang 2012);

auxiliary ***guo*** aspect article (Zhu 2014: 74)

例文：她去过中国。（彼女は中国に行ったことがある。）、

他吃过北京烤鸭。（彼は北京ダックを食べたことがある。）

ID55_列挙助詞“等/等等”：auxiliary ***deng*** (means “etc.”)　(Zhu 2014: 75)

例文：宋代著名诗人有陆游、王安石等。[44]

（宋代の有名な詩人には陸游・王安石などがいる。）

ID59_比況助詞：auxiliary (一样, 一般, 似的, etc.) used in the adjective comparative structures

例文：棉花白得像雪似的。[45]（綿は白くてまるで雪のようである。）

言語項目 ID： 60-62

「介詞」は Biber（2006: 184）の ‘Prepositional phrases’によるものである。英語では前置詞で始まる語句があるが、中国語においてこの概念に対応するのが介詞である。「接続詞」は黄・廖（2017 下冊: 27-28）によるものであり、NLPIR における接続詞のタグは ‘/cc’ また ‘/c’[46]である。「語気詞」は Zhu（2015: 76）の ‘interjection’という言語項目によるもので、タグが ‘/y’である。下記に示したのは、各言語項目に属する語彙の例である。

ID 61_介詞：在/p、跟/p、对/p、向/p

ID 62_接続詞：因为/c、可是/c、不但/c、而且/c

ID 63_語気詞：啊/y、呢/y、焉/y、呀/y

[44] この例文及び日本語の訳は「白水社 中国語辞典」の例文により引用したものである。URL: https://cjjc.weblio.jp/content/%E7%AD%89 〔最終アクセス日： 2022 年 10 月 16 日〕

[45] この例文及び日本語の訳は「白水社 中国語辞典」の例文により引用したものである。URL: https://cjjc.weblio.jp/content/%E4%BC%BC%E7%9A%84〔最終アクセス日： 2022 年 10 月 16 日〕

[46] NLPIR のタグセットにおいて、‘/cc’ は並列関係を表す接続詞のタグであり、 ‘/c’は並列関係以外の接続詞のタグである。

言語項目 ID： 63-66

「語詞」と「挿入語」は袁・李（2005）[47]によって整理された言語項目である。「省略語」と「ぼかし言葉」は Biber（2006: 184）における‘contractions’と‘adverbials: hedges’という 2 つの言語項目によるものである。「省略語」と「ぼかし言葉」の語彙リストを特定した際、凌（2000）の省略語、及び、董（2003）・方（2013）・洪（2015）[48]のぼかし言葉に関する研究を参考にした。下記に示したのは、各言語項目に属する語彙の一部（30 語）である。

ID 63_口語詞：老妈、老爸、信你个鬼、半吊子、兔崽子、挨班儿、挨个儿、挨肩儿、挨宰、矮半截儿、八九不离十、八字没一撇、巴不得、扒拉、把门、把式、爸爸、白搭、白玩儿、白眼儿狼、半熟脸儿、半中腰、包圆儿、背黑锅、背静、背气、奔命、绷脸、够呛、没辙

ID 64_挿入語：你想、你看、客观地说、保守地说、不客气地说、概括地说、坦率地说、实事求是地讲、说句不客气的话、说句不好听的话、说句不留情面的话、严格来讲、坦白地说、实话说、在我看来、对我来说、我觉得、没想到、你看看、您看看、你瞧、你瞧瞧、你瞧瞧你、你说、你说说、你还别说、你还真别说、用你的话说、不是我说你、你想想

ID 65_省略語：冬奥会、阿里、博导、家教、世博会、世贸组织、空勤、文史、医嘱、足联、足协、作协、北体、世卫组织、古罗马、阪大、东大、北大、北工、北航、边检、博导、车险、冬运会、二战、冬泳、房管、体训、阿里、奥理会

ID 66_ぼかし言葉：很、很多、很少、左右、至少、至多、约摸、有些、有点、一些、一点、一部分、许多、相当、未免、太、似乎、时常、少数、稍微、上下、若干、频频、偶尔、某种程度上、绝大多数、经常、尽量、几乎、极少数

---

[47] 袁晖・李熙宗（2005）.《汉语语体概论》.北京: 商务印书馆.

[48] 凌远征（2000）.《现代汉语缩略语》. 语文出版社.
方喻（2013）. 关于外国学生的汉语模糊限制语习得考察及教学对策（Master's thesis, 复旦大学）.
董娜（2003）. 模糊限制语的界定及分类. 北京第二外国语学院学报, (4), 28-34.
洪水英（2015）. 模糊限制语与汉语议论文语言分析.齐齐哈尔大学学报: 哲学社会科学版, (2), 127-130.

言語項目 ID： 67-69

「韻律単音節語」、「韻律双音節語」と「古語詞」は馮勝利（2006）によるものである。「韻律単音節語」（嵌偶单音词: monosyllabic word used in a disyllabic template）は「単+単」の構成に厳密に従うと述べられている（馮・王・黄 2008）[49]。その定義について、馮（2006）の『汉语书面用语初编(Expressions of Written Chinese)』（pp.2）では、下記のように述べている。

馮勝利（2006: 2）:

> “嵌偶单音词”是指当代书面正式语体中必须通过韵律组合才能独立使用的单音节文言词（亦即 a classical monosyllabic word must be formed as a prosodic word in order to be used in modern written Chinese）；换言之，嵌偶单音词要求和另一个单音词（或者也是嵌偶单音词，或者不是）合成一个“韵律词(Prosodic word)”才能合法出现。
>
> 訳：「韻律単音節語」[50]とは、現代の正式な書面語においては（他の単音節語との――*引用者注*）韻律結合（言葉の響きのよくなるような組み合わせ）を経ないかぎり使用できない文言単音節語(a classical monosyllabic word must be formed as a prosodic word in order to be used in modern written Chinese)である。いいかえるなら韻律単音節語は、他の単音節語（同じく韻律語であってもなくてもよい）と結合して「韻律語（Prosodic word）」を構成しない限りは文法にかなった使用が不可能である。

つまり、文言である韻律単音節語は文章に現れるために、必ず他の単音節語との結合で 2 音節化にしなければならない。例えば、下の表に示したように、“暗”と“返”は単独で使用することができず、必ず他の単音節語と結合（例、“<u>暗</u>查”、“<u>返</u>京”）しないと使用できない。表 3.3 は馮・王・黄（2008）の韻律単音節語の例に基づいて作成されたものである。

---

[49] 冯胜利・王洁・黄梅(2008)：嵌偶词严格地遵循着“单+单”的组构方式。

[50] “嵌偶单音词”は、「韻律単音節語」とも「韻律語の形成要素としての単音節語」とも訳すことができる。

表 3.3　「韻律単音節語」の例

韻律単音節語	单+单→双	誤例（单+双）
暗	暗查，暗送，暗想	暗检查，暗赠送，暗思考
备	备尝，备感，备受	备尝试，备感到，备经受
餐	订餐，送餐，三餐	预订餐，外送餐，三顿餐
错	错砍，错杀，错认	错砍伐，错杀死，错辨认
返	返京，返美，返校	返北京，返美国，返学校
广	广传，广交，广寻	广传播，广结交，广寻找
景	观景，选景，雪景	观赏景，选择景，白雪景
力	力保，力拒，力劝	力保证，力拒绝，力劝说
享	同享，独享，尽享	共同享，单独享，尽情享
资	闲资，添资，敛资	闲余资，添加资，收敛资

注：冯胜利・王洁・黄梅（2008）の「表 1 嵌偶词举例」に基づいて作成

また、「韻律双音節語」（合偶双音词: Disyllabic Words Combined with Another Disyllabic Word）の定義について、馮・王・黄（2008）は下記のように述べている。

馮・王・黄（2008）：

现代汉语书面语还自身发展出一批双音词，这类双音词文言中没有，口语中罕用，只在现代汉语书面语中出现，且必须和另一个双音词组成“双+双”的韵律模块后才能合法出现，这种双配双的双音词称之为“合偶双音词”。

訳：現代中国語の書き言葉は（韻律単音節語の他に）さらに一群の双音節語を、それ自体から発展させてきた。この双音節語は文語には見られず、口語でも稀にしか用いられず、また、別の双音節語と「双+双」の韻律的な組み合わせを作らない限りは、規範にかなった形で用いることができない。別の双音節語と組み合わせて用いるこのような双音節語を「韻律双音節語」と呼ぶ。

表 3.4 は馮・王・黄（2008）の韻律双音節語の例に基づいて作成されたものである。

表 3.4　「韻律双音節語」の例

韻律双音節語	双+双	誤例（双+单）
保卫	保卫人民，保卫祖国，保卫家乡	保卫人，保卫国，保卫家
从事	从事教学，从事写作，从事翻译	从事教，从事写，从事译
光临	光临寒舍，光临我校，光临大会	光临家，光临校，光临会
建筑	建筑桥梁，建筑公路，建筑房屋	建筑桥，建筑路，建筑房
合法	合法居住，合法买卖，合法分配	合法住，合法买，合法分
宏伟	宏伟气势，宏伟宫殿，宏伟蓝图	宏伟气，宏伟宫，宏伟图
极为	极为严格，极为深刻，极为灵验	极为严，极为深，极为灵
加以	加以评论，加以改正，加以补充	加以评，加以改，加以补
进行	进行调查，进行讨论，进行批判	进行查，进行论，进行批
消耗	消耗能源，消耗体力，消耗时间	消耗能，消耗力，消耗时

注：冯胜利・王洁・黄梅（2008）の「表 2 合偶词举例」に基づいて作成

本研究の「韻律单音節語」、「韻律双音節語」と「古語詞」の言語項目に属する語彙や表現の一部（30 語）は以下の通りである。

ID 67_韻律单音節語：案、暗、饱、暴、倍、备、奔、鼻、避、遍、别、鬓、并、捕、彩、餐、残、惨、畅、晨、尘、诚、承、重、出、窗、春、翠、错、挫

ID 68_韻律双音節語：安定、安装、办理、保持、保留、保卫、保障、报道、暴露、爆发、被迫、必然、必修、必要、避免、编制、变动、变革、辩论、表达、表示、表演、并肩、补习、不断、不时、不住、布置、采取、采用

ID 69_古語詞：言必称、的状态、称之、之势、出入于、之位置、不次于、从属于、得自于、得力于、大大发挥、发轫于、凡此皆、凡此种种皆、所能及、庶可、之故、工于、广为、跻于、之列、不日即、之大成、以俱来、见托于、见之于、见胜、混同于、归之于、而论

言語項目 ID： 70-73

低難度語彙・中難度語彙・高難度語彙・非 HSK 語彙は HSK 語彙リスト[51]に基づいて選定した言語項目である。HSK 語彙リスト（計 5000 語）の構成は下記の通りである。

図 3.4　HSK の各レベルに対応する語彙量

本研究の語彙難度の区分について、低難度語彙は HSK1−2 級の語彙（計 300 語）、中難度語彙は HSK3−4 級の語彙（計 900 語）、高難度語彙は HSK5−6 級の語彙（計 3800 語）の集合であり、非 HSK 語彙は HSK 語彙リスト（5000 語）に収録されていない語彙を指すことである。低難度語彙、中難度語彙、高難度語彙の単語リストにおける上位 50 語彙は以下の通りであるが。HSK1−6 級の各レベルに対応する単語リストは、本稿末の付録 III を参照されたい。

ID70_低難度語彙：爱、八、爸爸、杯子、北京、本、不、不客气、菜、茶、吃、出租车、打电话、大、的、点、电脑、电视、电影、东西、都、读、对不起、多、多少、儿子、二、饭店、飞机、分钟、高兴、个、工作、狗、汉语、好、号、喝、和、很、后面、回、会、几、家、叫、今天、九、开、看

[51] HSK は（中国語名: 漢語水平考試）は中国の教育部が認定する国際的な中国語の語学検定試験である。HSK の公式サイトには、1 級（最低級）~ 6 級（最高級）で使用される語彙（計 5000 語）のリストは《新汉语水平考试(HSK)词汇(2012 年修订版)》の名称で配布されている。URL: https://www.chinesetest.cn/godownload.do〔最終アクセス日:2022 年 5 月 20 日〕

ID71_中難度語彙：阿姨、啊、矮、爱好、安静、把、班、搬、办法、办公室、半、帮忙、包、饱、北方、被、鼻子、比较、比赛、笔记本、必须、变化、别人、冰箱、不但、而且、菜单、参加、草、层、差、超市、衬衫、成绩、城市、迟到、除了、船、春、词典、聪明、打扫、打算、带、担心、蛋糕、当然、地、灯、地方

ID72_高難度語彙：哎、唉、爱护、爱惜、爱心、安慰、安装、岸、暗、熬夜、把握、摆、办理、傍晚、包裹、包含、包括、薄、宝贝、宝贵、保持、保存、保留、保险、报到、报道、报告、报社、抱怨、背、悲观、背景、被子、本科、本领、本质、比例、彼此、必然、必要、毕竟、避免、编辑、鞭炮、便、辩论、标点、标志、表达、表面

言語項目 ID： 74-78

中国語の文法において、“把”構文（中国語：“把”字句）と受身文（中国語：被动句）は重要な存在である。本研究で設けた「“把”構文」という言語項目は Zhu(2015: 74)の‘preposition Ba’によるものであり、「受身文、動作主なしの受身文、過去を表す構文、動作の進行を表す構文」は Biber(2006: 184)における‘passives: by-phrase’、‘passives: agentless’、‘verbs: past tense’、‘verbs: progressive’を参考にして設けた言語項目である。下記に示したのは、各言語項目に対応する例文である。

ID 74_“把”構文：把/pba

例文(タグ付け)：老师/n **把/pba** 他/rr 叫/v 走/v 了/ule

ID 75_受身文：被/pbei

例文(タグ付け)：小王/nr **被/pbei 老师**/n 骂/v 了/ule

ID 76_動作主なしの受身文：

例文(タグ付け)：小王/nr 被/pbei 骂/v 了/ule

ID 77_過去を表す構文：已经/曾经/已/曾/刚/刚刚/方才 + V、V + 过/了

例文(タグ付け)：我/rr 已经/d 吃/v **过/guo** 午饭/n 了/ule

ID 78_動作の進行を表す構文：正/在/正在+V、V+之时、当/正当+V+时、V+着

例文（タグ付け）：他/rr 正在/d 看/v 书/n

言語項目 ID： 79-88

中国語には英語のような節（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・廖（2017: 129-139）で考察されている 10 種類の複文（言語項目の ID: 79-88）は分析に含まれている。下記は各複文に属するパターンや語彙の一部を示している。

ID 79_並列複文： 既 A, 又（也）B；一边 A, 一边 B；不是 A, 而是 B；有时 A, 有时 B；一方面 A, 一方面 B；不是 A, 而是 B；并非 A, 而是 B；是 A, 不是 B

ID 80_連続複文： 首先 A, 然后 B；先 A, 再 B；刚 A, 就 B；一 A, 就 B

ID 81_解説複文： 即；就是说

ID 82_選択複文： 或者 A, 或者 B；是 A, 还是 B；不是 A, 就是 B；要么 A, 要么 B

要不 A, 要不 B；与其 A, 不如 B；与其 A, 毋宁 B；宁可 A, 也不 B

ID 83_累加複文： 不但 A, 而且 B；不仅 A, 还 B；尚且 A, 何况 B；别说 A, 就是 B 也

ID 84_条件複文： 只要 A, 就 B；一旦 A, 便 B；只有 A, 才 B；除非 A, 才 B

ID 85_仮定複文： 如果 A, 就 B；要是 A, 则 B；即使 A, 也 B；再 A, 也 B

ID 86_因果複文： 因为 A, 所以 B；由于 A, 于是 B；既然 A, 那么 B

ID 87_目的複文： 以便；为的是；以免；免得；省得；以防

ID 88_逆接複文： 虽然 A, 但是 B；尽管 A, 也 B；虽说 A, 然而 B；固然 A, 可是 B

言語項目 ID： 89-92

コーパス分析の基本項目として、語彙多様性、語彙密度、単語の平均長さ、平均文長という 4 種の言語項目を特定した。「語彙多様性」、「単語の平均長さ」と「平均文長」

は Zhu（2015）の “type token ratio” 、“mean word length” 、“mean sentence length” によるもの[52]で、「語彙密度」は本研究で新しく特定された言語項目である。

「語彙多様性」とは、テキストに使用されている語彙の豊富さの度合いを指し、最も一般的な計算方法は Type/Token（TTR: Type-Token Ratio）である（Van Hout & Vermeer 2007）。Zhu（2015）と Biber（2006: 184）は、“Type/Token”を語彙多様性の計算式として採用している。しかし、TTR がテキストの長さの影響を受けやすいため、長さ依存の問題に対する代替案として、Uber index（計算式：logTokens – logTypes）や Guiraud index（計算式：Type /√ Token）などが提案されている（Jarvis 2002）。なお、Guiraud index（以下は「Guiraud 値」とする）は、語彙多様性を測る最も信頼性の高い指標の一つと考えられている（Van Hout & Vermeer 2007）。Verspoor *et al*（2012）では、Guiraud 値が L2 学習者の作文における語彙の多様性の変化を、英語レベルの違いによって明確に区別できる非常に有用な指標であることを示した。そのため、本研究は Guiraud 値（計算式：Type /√ Token）を語彙多様性の計算方法として採用する。

「語彙密度」とは、テキストに含まれる内容語の割合である。テキストの語彙密度が高いほど、そのテキストに含まれる情報量は多くなると述べられている（黄・王 2020）。

「平均文長」と「単語の平均長さ」とはそれぞれ、センテンスと単語の長さの平均値を表す指標である。平均文長に関する研究の多くは、文体論や談話領域に焦点が当てられてきた。例えば、Morton（1965）は、ギリシャ語の散文の著者を推定するためにセンテンスの長さを利用した。Dzurjuk（2006）は、ドイツ人作家の作品を研究するために、センテンスの長さを文体測定の尺度に取り入れた。また、中国語の文の長さ（Chinese sentence length）に関する先行研究は、中国語の言語使用域（Chinese register）の予測との関連性を確立していると Hou *et al*（2019）は述べている。例えば、劉・胡（2011）は、センテンスの長さの分布は話し言葉のニューステキストと書き言葉のニュースを区別できると結論づけた。Wu（2005）は中国語のセンテンスの長さを文字数で定義し、この特徴を利用して中国語のニュースとフィクションを区別した。

Jiang（2013）は中国語で書かれた文の長さを、単語数より文字数で数えるべきだと述べている[53]。なぜなら、中国語の文章はスペースや単語の区切り記号（句読点を除く）を

[52] Biber（2006: 184）においても“type/token ratio”と “average word length”を設けている。

使わずに書かれるため、中国語の単語の境界が不明確だからであると述べられている（Jiang 2013）。そのため、本研究では、センテンスと単語の長さを文字数で測る。

これら 4 つの言語項目の計算式は下記の通りである。

ID 89_語彙多様性： Types / $\sqrt{\text{Tokens}}$

ID 90_語彙密度： content words / Tokens

ID 91_単語の平均長さ： characters / Tokens

ID 92_平均文長： characters / sentences

3.3 言語項目の集計

これら 92 の言語項目の集計は、主に Python を用いて作成したコンピュータプログラムによって行われた。下記では、言語項目の頻度集計に関するプログラミングの一部のコードを例として示している。

（1）Wordlist による頻度集計

92 種の言語項目において、一部の言語項目は単語リストで特定されたものである。

[53] Jiang (2013): One would tend to count *zi* (characters) in any written discourse in Chinese, not only because *zi* is the basic unit in Chinese language but also because the boundary of a Chinese syntactic word is rather unclear, since Chinese is written without using any spaces or other word delimiters (except for punctuation marks).

例文：

text = “我/rr 和/cc 她们/rr 是/vshi 在/p 十五/m 年/qt 前/f 上/vf 小学/n 的/ude1 时候/n 认识/v 的/ude1 。/wj 我们/rr 小学/n 规模/n 非常/d 小/a ，/wd 全校/n 儿童/n 是/vshi 七十/m 名/q 左右/m ，/wd 我们/rr 班/n 有/vyou 十二/m 个/q 同学/n 。/wj 虽然/c 跟/p 全班/n 同学/n 关系/n 真/d 好/a ，/wd 但是/c 我/rr 跟/p 奥原/nrj 和/cc 小村/nrj 特别/d 友好/a ，/wd 因为/c 我们/rr 学号/n 很/d 近/a 。/wj 我们/rr 每天/r 下学/v 后/f 一起/d 去/vf 玩儿/v ，/wd 去/vf 乒乓球/n 俱乐部/n ，/wd 上/vf 初中/n 以后/f 也/d 一起/d 参加/v 乒乓球/n 社团/n 了/y 。/wj 我们/rr 是/vshi 所谓/v 的/ude1 竹马/n 之/uzhi 交/ng ，/wd 一起/d 过/vf 了/ule 比/p 家人/n 更/d 长/a 的/ude1 时间/n 。/wj 虽然/c 我们/rr 进入/v 了/ule 别的/rzv 高中/n ，/wd 但是/c 有时/d 和/cc 她们/rr 见面/vi 一起/s 去/vf 吃饭/vi ，/wd 现在/t 还/d 和/cc 她们/rr 保持/v 联系/vn 。/wj”

例えば、上記のタグ付けの例文における「ID11_高頻度動詞」の出現頻度（粗頻度：raw frequency）を集計するための主なコードは下記の通りである。「#」の後ろに示したのは、各行のコードに関する注釈である。

```
---------------------------------------------------------------------------------------------------------
# split() メソッドを用い、空白（スペース）でテキストを分割する。
Wordlist = text.split()

# re.search メソッドを用い、Wordlist における動詞を抽出する。動詞のタグは「/v」である。
動詞_tokens = [token for token in Wordlist if re.search("/v",
                token)]

# split("/")[0]メソッドを用い、動詞_tokens における動詞のタグを除き、文字の部分を残す。
動詞_token = [token.split("/")[0] for token in 動詞_tokens]

# 高頻度動詞の wordlist を読み込む。
高頻度動詞_wordlist = [line.strip() for line in open(
            '高頻度動詞_wordlist.txt', 'r', encoding='utf-8').readlines()]

# テキストにおいて、高頻度動詞に属する動詞の頻度を集計する。
高頻度動詞_frequency = len([token for token in 動詞_token if token
                in 高頻度動詞_wordlist])
---------------------------------------------------------------------------------------------------------
```

(2) 単語のタグによる頻度集計

言語項目の特徴によって、単語のタグに基づいて頻度集計される言語項目は少なくない。例えば、上記のタグ付けの例文における「ID 38_一人称単数代名詞“我”」、「ID 39_一人称複数代名詞“我们”」、「ID52_構造助詞“的”」や副詞、接続詞の出現頻度を集計するための主なコードは下記の通りである。

```
---------------------------------------------------------------------------------------------------------
# split() メソッドを用い、空白（スペース）でテキストを分割する。
Wordlist = text.split()

# Wordlist における一人称単数代名詞“我”の頻度を集計する。
一人称単数代名詞_我_frequency = len([token for token in Wordlist if
                token in ["我/rr"]])

# Wordlist における一人称複数代名詞“我们”の頻度を集計する。
一人称複数代名詞_我们_frequency = len([token for token in Wordlist if
                token in  ["我们/rr "]])

# Wordlist における構造助詞“的”の頻度を集計する。
構造助詞_的_frequency = len([token for token in Wordlist if token in
                ["的/ude1"]])

# Wordlist における副詞の頻度を集計する。副詞のタグは「/d」である。
副 詞 _frequency = len([token for token in Wordlist if
                token.split("/")[-1] in ["d"]])

# Wordlist における接続詞の頻度を集計する。接続詞のタグは「/c」と「/cc」である。
接 続 詞 _frequency = len([token for token in Wordlist if
                token.split("/")[-1] in [ "cc","c"]])
---------------------------------------------------------------------------------------------------------
```

これら 92 種の言語項目を集計するために、本研究では、コンピュータプログラムのシステムを開発した。このプログラムは 2022 年 9 月にオンラインで一般公開された。

URL：https://count-linguistic-features.herokuapp.com/

上記の URL をクリック[54]すると、下記のような言語項目の頻度集計画面が表示される（図 3.5）。

[54] 将来はこのプログラムを変更や編集する可能性がある。

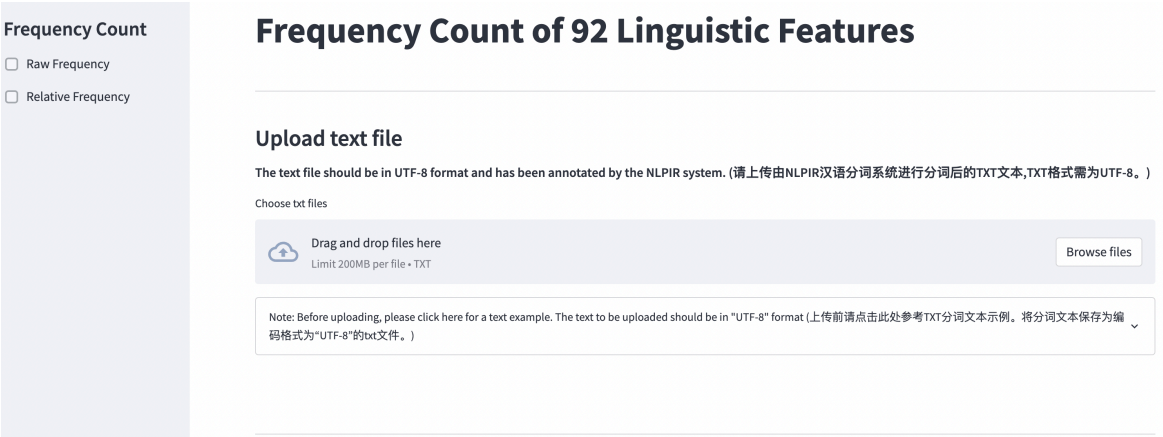

図 3.5　言語項目の頻度集計プログラム

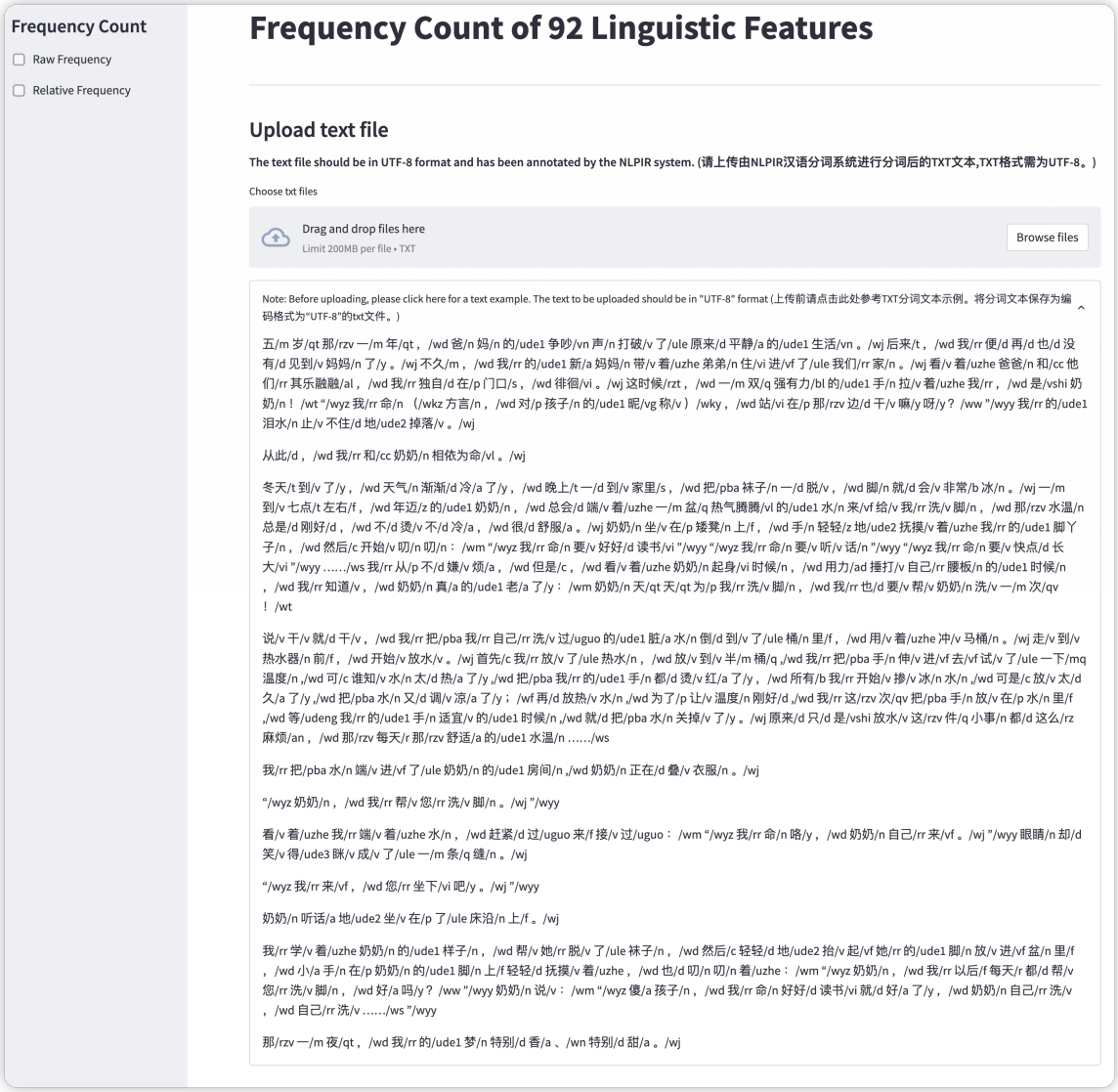

図 3.6　TXT ファイルの例

図 3.5 に示した「Browse files」から TXT ファイル（例：図 3.6 のようなタグを付けたテキスト）をアップロードし、左側の「Raw Frequency」と「Relative Frequency」にチェックを入れると、92 種の言語項目の粗頻度と 1,000 字あたりの相対頻度を自動的に集計することができる（図 3.7）。

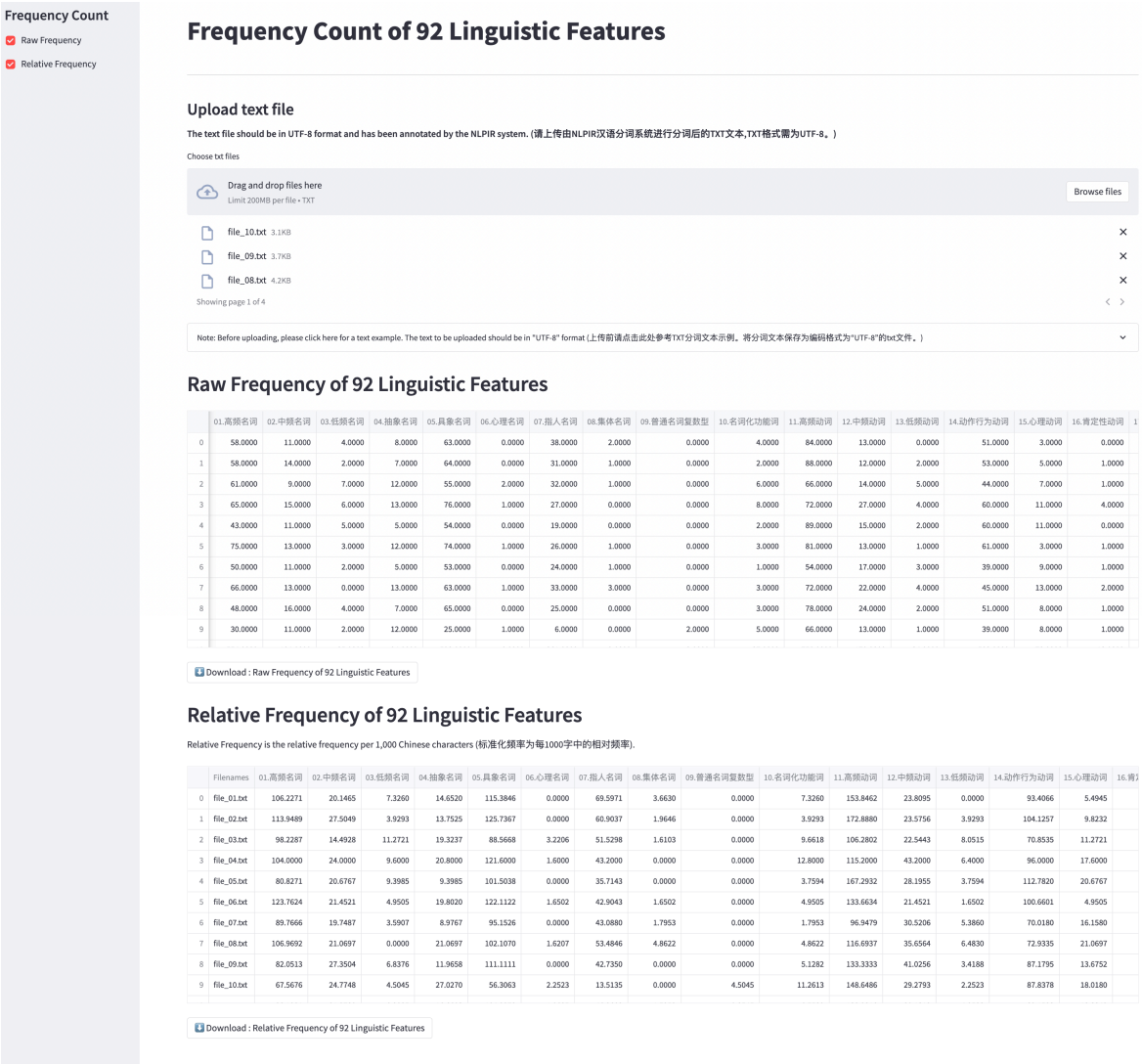

図 3.7　頻度集計の結果

1,000 字あたりの相対頻度について、簡単に説明する。例えば、「file_04.txt」ファイルの総字数（句読点は字数にカウントされない）は 625 字であり、そのファイルにおける「ID 01_高頻度名詞」・「ID 02_中頻度名詞」・「ID 03_低頻度名詞」の粗頻度がそ

れぞれ 65 回、 15 回、 6 回である場合、これら 3 つの言語項目の 1,000 字あたりの相対頻度は下記の通りである。

ID 01_高頻度名詞　　65 ÷ 625 × 1000 = 104

ID 02_中頻度名詞　　15 ÷ 625 × 1000 = 24

ID 03_低頻度名詞　　6 ÷ 625 × 1000 = 9.6

ここで強調したい点がある。92 種の言語項目のうち 4 つの言語項目の値（ID 89~92：語彙多様性、語彙密度、単語の平均長さ、平均文長）は、1,000 字あたりの相対頻度に従って調整することはせず、3.2 節の最後に示した計算式に従って自動的に算出することである。

ID 89_語彙多様性： Types／√Tokens

ID 90_語彙密度： content words / Tokens

ID 91_単語の平均長さ： characters / Tokens

ID 92_平均文長： characters / sentences

図 3.7 に示した「⬇️ Download」ボタンをクリックすると、頻度集計の粗頻度と相対頻度の結果を CSV ファイルとしてダウンロードすることが可能である。頻度集計で得られた相対頻度表は、様々な統計分析に利用することができる。本研究は、得られた相対頻度表を用いて因子分析を行った（因子分析の詳細は第 5 章を参照されたい）。

第 4 章　コーパスの構築

本章は、中国語母語話者と日本人中国語学習者が書いた中国語作文のコーパス構築について述べる。第 1 節では、作文テキストの収集・分類及びコーパスの内訳を明記する。第 2 節では、コーパスの検索機能、及び実際に検索した場合の手順と結果を示す。

4.1 作文コーパスの構築

表 4.1 はコーパスの基本情報であり、コーパスのサイズと、各サブコーパスの総字数及び総語数（Tokens）の情報を示している。

表 4.1　コーパスの基本情報[55]

Corpus	Sub_corpus	N	Tokens		Characters	
			総語数	平均	総字数	平均
中国語母語話者	CHN_人物描写	150	70,984	473.23	115,906	772.71
	CHN_出来事の叙述	150	68,608	457.39	113,738	758.25
	CHN_意見文	150	73,023	486.82	127,755	851.70
	CHN_手紙	150	75,164	501.09	124,995	833.30
	CHN_日記	150	65,724	438.16	109,785	731.90
	計	**750**	**353,503**	**471.34**	**592,179**	**789.57**
日本人中国語学習者	JP2_人物描写	183	46,343	253.24	79,956	436.92
	JP2_出来事の叙述	43	10,764	250.33	18,094	420.79
	JP2_手紙	82	22,985	280.30	40,066	488.61
	JP2_日記	163	41,041	251.79	71,323	437.56
	JP3_人物描写	111	41,801	376.59	70,459	634.77
	JP3_出来事の叙述	61	26,622	436.43	47,654	781.21
	JP3_意見文	57	21,704	380.77	38,759	679.98
	計	**700**	**211,260**	**301.80**	**366,311**	**523.30**
	合 計	**1,450**	**564,763**	**389.49**	**958,490**	**661.03**

[55] N はテキスト数、総字数は句読点を含む字数、総語数は句読点を含まない単語数である。

分析対象とするテキストは中国語を専攻としている日本人大学 2・3 年生によって書かれた中国語作文である。学習者の作文は、5 つの作文ジャンル（手紙・日記・出来事の叙述・人物描写・意見文）からなる 700 テキストであり、学習者の中国語レベル[56]と受講するコースによって、中級作文コーパス（以下は「JP2」とする）と上級作文コーパス（以下は「JP3」とする）に分類される。また、学習者の作文における言語特徴と比較対照するために、本研究では“zuowen 网” （中国語作文の WEB サイト）から中国語母語話者が書いた 5 ジャンル[57]の作文を収集し、構築した母語話者作文コーパス（以下は「CHN」とする）を比較対象とする。

4.2 コーパスの検索機能とその応用

上記で収集してきた作文テキストをオンラインコーパス（Online Corpus）として構築した。2022 年 3 月時点で設けた検索機能は、主にキーワード（Keyword）によるセンテンス（Sentence）の検索、品詞（POS）の検索、ワードリスト（Wordlist）の検索、Concordance である。本研究で構築したオンラインコーパスは、2022 年 9 月に以下の URL で一般公開された[58]。

URL：https://xuqin-composition-corpus.herokuapp.com/

4.2.1 Keyword による Sentence の検索

キーワードによるセンテンスの検索（Keywords in Sentence）とは、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。

上に述べたような「センテンス」はフルセンテンス（full sentence）であり、主に中国語の句点「。」、疑問符・クエスチョンマーク「?」、感嘆符「！」、省略記号「……」

---

[56] 中級（JP2）は「中級中国語作文」の授業を受講し、かつ中国語レベルが HSK3-4 級に相当する。上級（JP3）は「上級中国語作文」の授業を受講し、かつ中国語レベルが HSK5-6 級に相当する。また、 HSK（漢語水平考試）は中国教育部が認定する国際的な中国語の語学検定試験であり、 1 級から 6 級まであり、HSK 6 級の難度が一番高い。

[57] 5 つの作文ジャンルは、“写人”（人物描写）、“叙事”（出来事の記述）、“日记”（日記）、“书信”（手紙）、“议论文”（ 意見文）である。“zuowen 网”（中国語作文 WEB サイト）：http://www.zuowen.com/gaozhong/〔最終アクセス日: 2022 年 5 月 20 日〕

[58] 今後、作文テキストや新しい検索機能の追加など、コーパスに手を加える可能性がある。

という 4 つの記号（標点符号）で識別される。キーワードによるセンテンスの検索手順は下記の通りである。

（1）左側にある「Sentence Search」のチェックボックスにチェックを入れる（図 4.1）。

（2）サブコーパス「JP2」・「JP3」・「CHN」のマルチセレクトボックス（multi-select box、複数選択可）から作文のジャンルを選択する。ここでは「JP2」・「JP3」・「CHN」サブコーパスにおける“叙事”（出来事の叙述）をジャンルの例として選択した（図 4.1 を参照）。

図 4.1　作文ジャンルの選択とキーワードの入力

（3）キーワードは 5 つまで入力することが可能である。「Keywords in Sentence」の下側にキーワードを入力し、「Search」ボタンを押すことにより、検索結果が表示される（図 4.2 ）。ここでは、“我”（私）、“非常”（非常に）、“开心”（うれしい）の 3 つの単語をキーワードの例として入力して検索した。

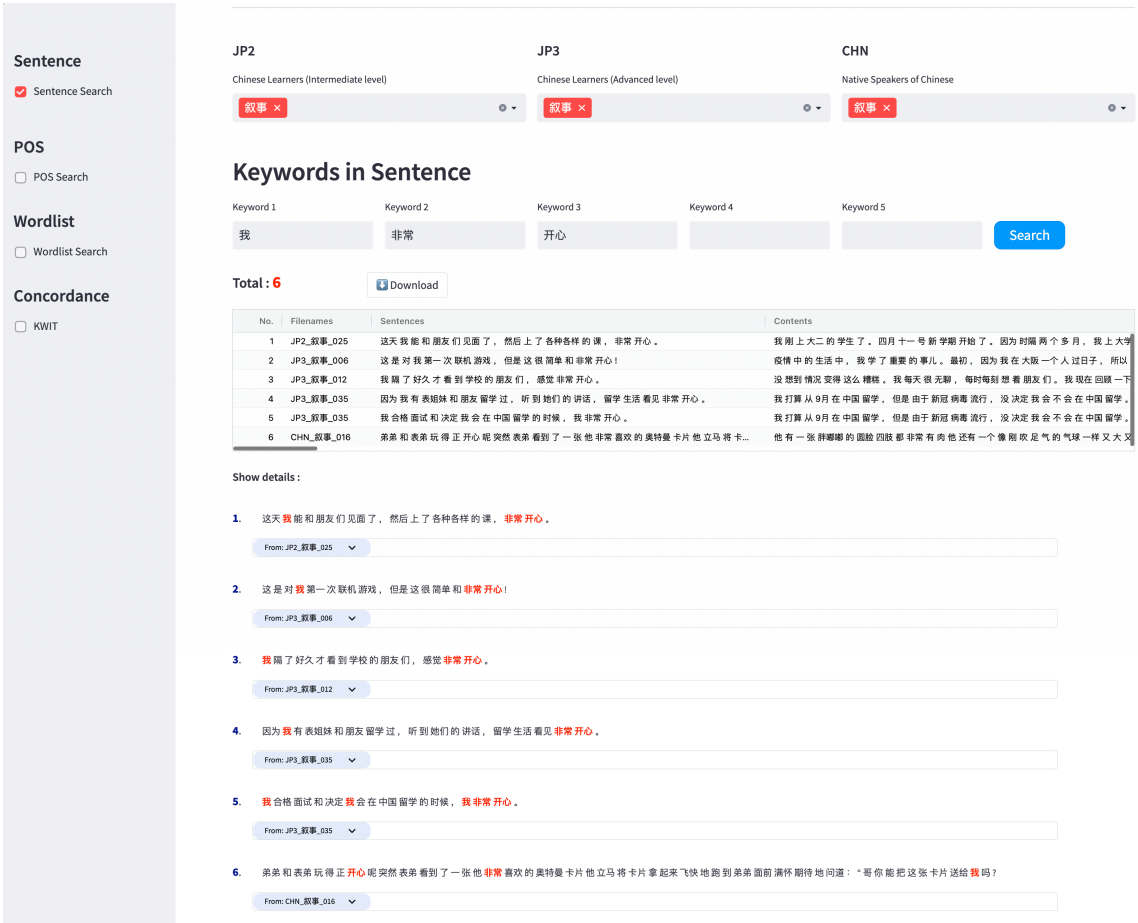

図 4.2　キーワードによるセンテンス検索の結果

検索結果（図 4.2）を見ると、「Total: 6」で示したように、“我”・“非常”・“开心”を含めるセンテンスが 6 つであることが分かった。また、検索した結果は AgGrid セルの形式で示されている。AgGrid セルにおいて、1 列目は「No.」（センテンスの Id）、2 列目は「Filenames」（センテンスが対応するテキストの名前）、3 列目は「Sentences」（検索結果としてのキーワードが含まれるセンテンス）、4 列目は「Contents」（センテンスが対応するテキスト全体の内容）である。AgGrid セルの上部にある「Download」のボタンを押すと、検索の結果を CSV ファイルとしてダウンロードすることができる。

AgGrid セルの下方（Show details）に示したのは、具体的なセンテンスの情報である。センテンスにおけるキーワードは赤色で示されており、センテンスの出典も表示されている。例えば、検索結果としての 1 つ目のセンテンスの出典は「From: JP2_叙事_025」で表示されているように、1 つ目の検索結果は「JP2_叙事_025」から抽出されたセンテ

ンスであることを意味している。言い換えれば、「JP2_叙事」（JP2_出来事の叙述）のサブコーパスにおいて、25 番目のテキストには、入力したキーワードを含むセンテンスがあることが分かる。

最後に、図 4.3 に示すように、センテンスの出典をクリックすると、テキストの全体の内容を表示することができ、キーワードを含めるセンテンスが黄色でハイライトされている。「From: JP2_叙事_025」と「From: JP3_叙事_006」をクリックした結果例を下記のように示している。

図 4.3　センテンスに対応するテキスト内容

4.2.2 品詞の検索

中国語テキストからある種類の品詞、例えば、名詞・動詞・形容詞・副詞などを抽出したい場合、中国語の単語分割と附与されたタグに基づいて実現できる。品詞（POS）

の検索という機能では、あるジャンルにおける特定の品詞を抽出して、その品詞に対応する単語の粗頻度と相対頻度（1,000 字あたりの出現頻度）を自動的に示すことが可能である。

品詞（POS）による検索手順は下記の通りである。

（1）左側に示す「POS Search」のチェックボックスにチェックを入れる（図 4.4）。

（2）「POS Search」の下部にある「Select a category:」のセレクトボックス（select box）から 1 種の品詞を選択する。ここでは“连词”（接続詞）を例として選択した（図 4.4）。

（3）「JP2」・「JP3」・「CHN」のマルチセレクトボックス（multi-select box、複数選択可）から作文のジャンルを選択する。ここでは「JP2」・「JP3」・「CHN」コーパスにおける“叙事”（出来事の叙述）をジャンルの例として選択した（図 4.4）。

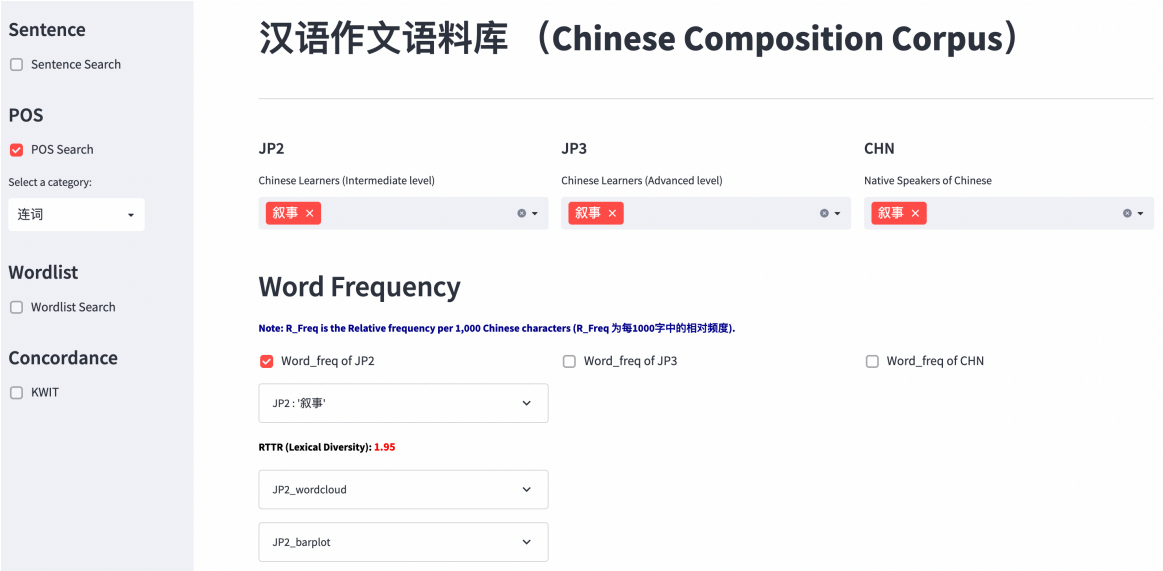

図 4.4　品詞類と作文ジャンルの選択

（4）「Word Frequency」の下部にあるチェックボックスにチェックを入れる。ここでは「Show word_freq of JP2」を例としてのチェックボックスにチェックを入れた。チェックを入れると、「JP2_叙事」（JP2_出来事の叙述）というジャンルにおける“连词”（接続詞）に関する語彙多様性（RTTR）が自動的に表示される（図 4.4）。「Word_freq of JP2」・「Word_freq of JP3」・「Word_freq of CHN」のチェックボッ

クスにチェックを入れることにより、「JP2」・「JP3」・「CHN」間の比較ができる（図 4.5）。

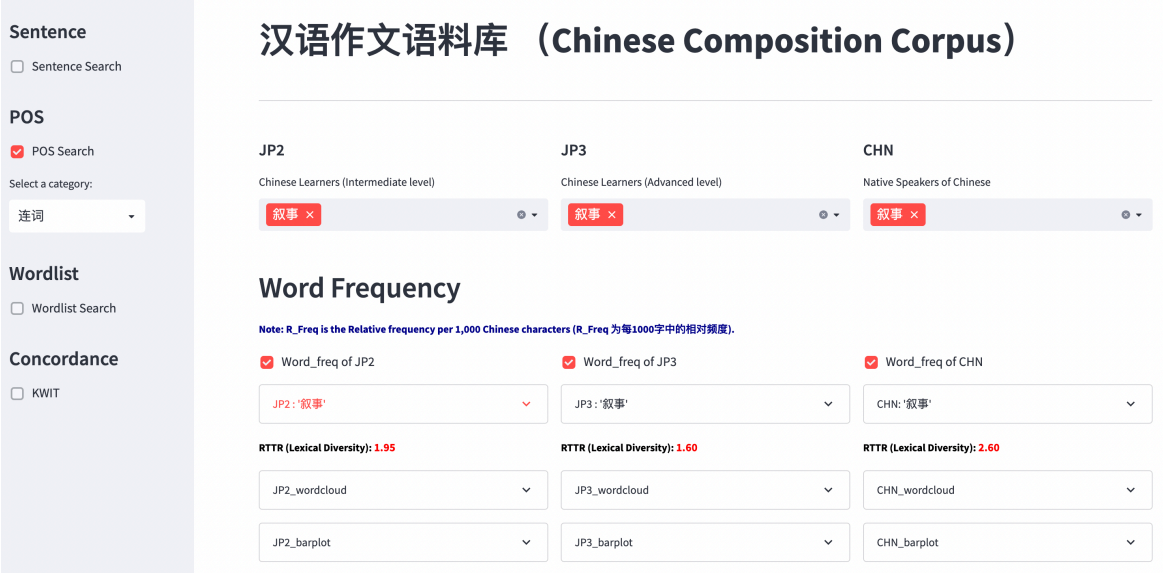

図 4.5　グループ間の比較

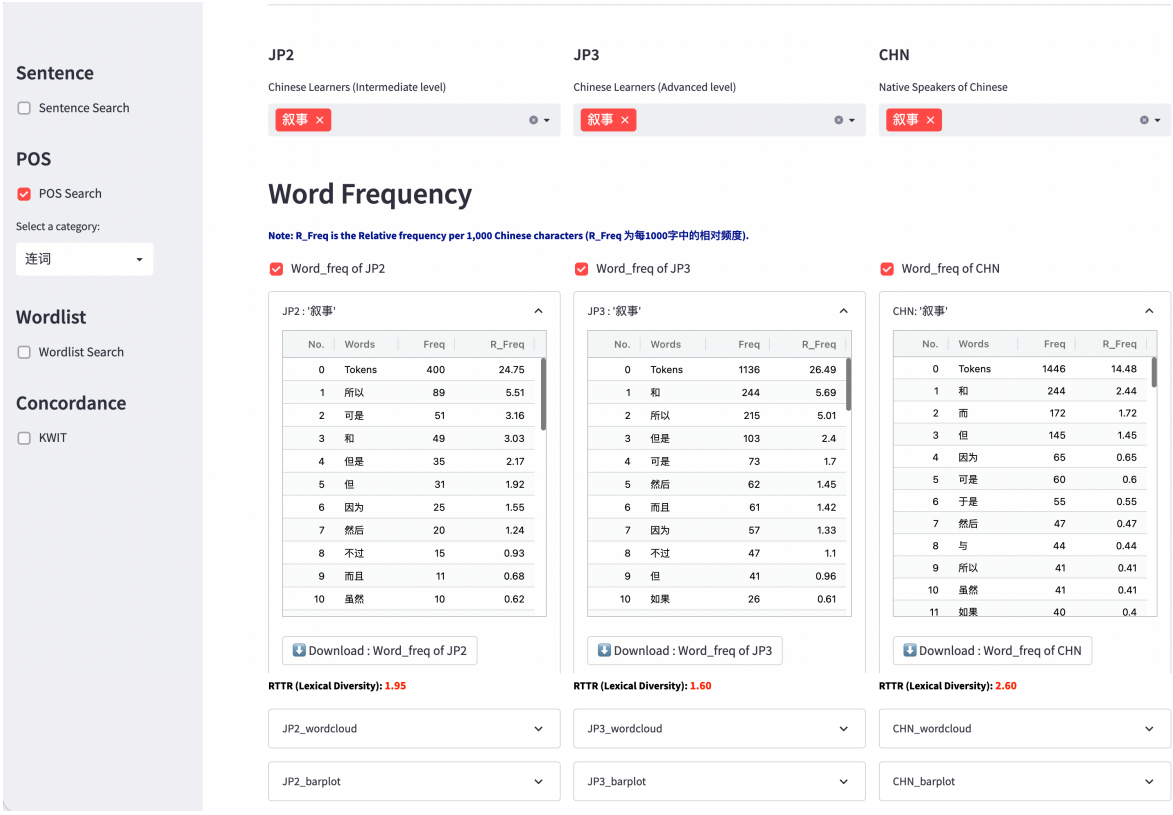

図 4.6　単語の表示

（5）「Word_freq of xx」の下にあるエキスパンダ（expander）を押すと、エキスパンダ領域が展開され、指定された品詞に属する単語を降順に集計することができる。例えば、図 4.6 に示したように、3 つのジャンルにおいて、“连词”（接続詞）に属する単語を降順に示している。セルの 1 行目はすべての単語（Tokens）の粗頻度（Freq）と相対頻度（R_Freq）であり、2 行目からは各単語の粗頻度と相対頻度である。

図 4.6 に見られるように、「JP2_叙事」・「JP3_叙事」・「CHN_叙事」という 3 つのジャンルにおいて、“连词”（接続詞）の粗頻度（或いは、総頻度・総語数）はそれぞれ 400 回・1136 回・1446 回であり、1,000 字あたりの相対頻度はそれぞれ 24.75 回・26.49 回・14.48 回である。3 つのジャンルにおける上位 5 語の接続詞に関しては、「JP2_叙事」は“所以”・“可是”・“和”・“但是”・“但”、「JP3_叙事」は“和”・“所以”・“但是”・“可是”・“然后”、「CHN_叙事」は“和”・“而”・“但”・“因为”・“可是”であることも確認された（図 4.6）。

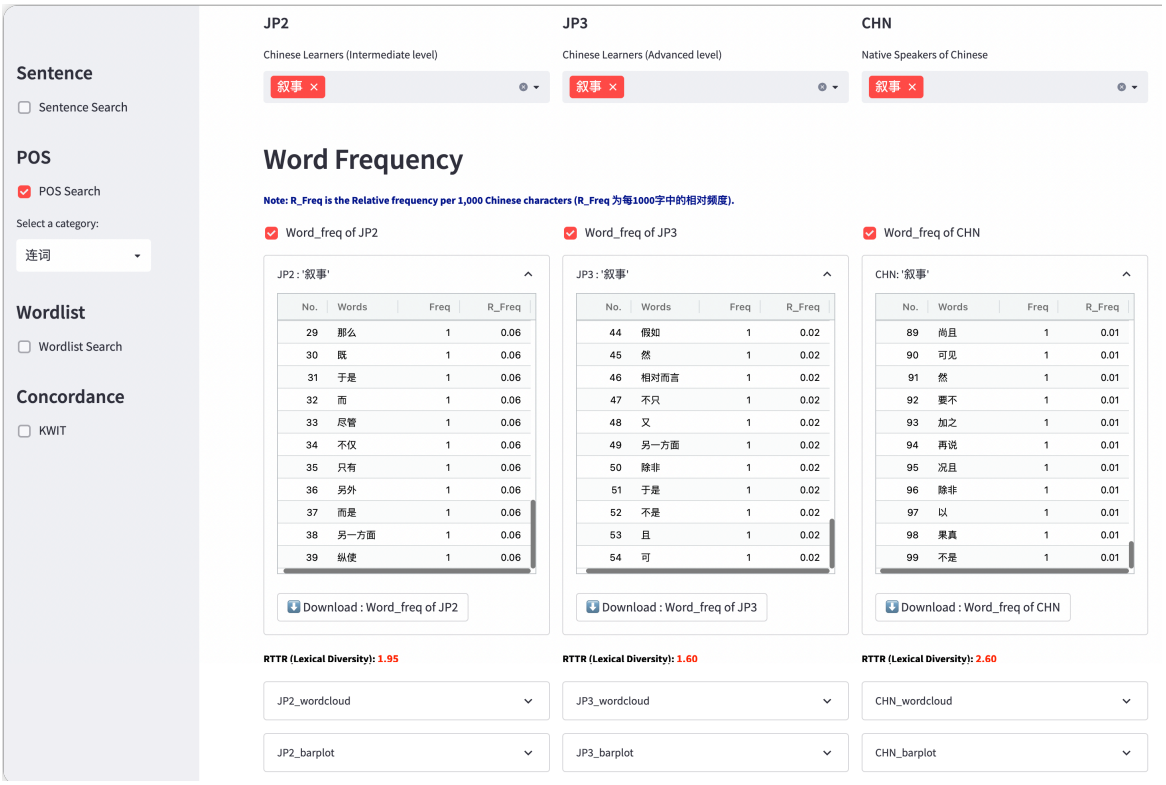

図 4.7　スクロールバーのドロップダウン

各セルの右横にある縦スクロールバーをドロップダウンすると、3 つのジャンルにおける接続詞の単語は合計 39 種・54 種・99 種であることが分かった（図 4.7）。各セルの下部には、各ジャンルに対応する語彙多様性（RTTR）[59]の値を示している。「JP2_叙事」と ・「JP3_叙事」において接続詞の相対頻度（約 25 回、26 回）は 「CHN_叙事」（約 14 回）の約 2 倍ほどであるが、接続詞の語彙多様性（RTTR）は「CHN_叙事」より低いことが分かった。

（6） 「RTTR（Lexical Diversity）: xx」の下にある wordcloud のエキスパンダ（expander）領域を押すと、エキスパンダ領域が展開され、指定した品詞種類に属する単語を wordcloud で可視化することができる。また、「wordcloud」の下にある bar plot のエキスパンダ（expander）領域にある「＋」を押すと、エキスパンダ領域が展開され、指定された品詞種類に属する単語の上位 20 語を Bar plot で表示できる（図 4.8）。

図 4.8　Word cloud と bar plot の表示

（7）各 bar plot の右にある矢印（View fullscreen）をクリックすると、bar plot を拡大して表示することができる（図 4.9）。

[59] 語彙多様性はテキストにおける語彙使用の豊富さを測る指標である。本研究は Guiraud が提唱した RTTR（計算方法：Type /√ Token ）を語彙多様性 の指標として採用する。

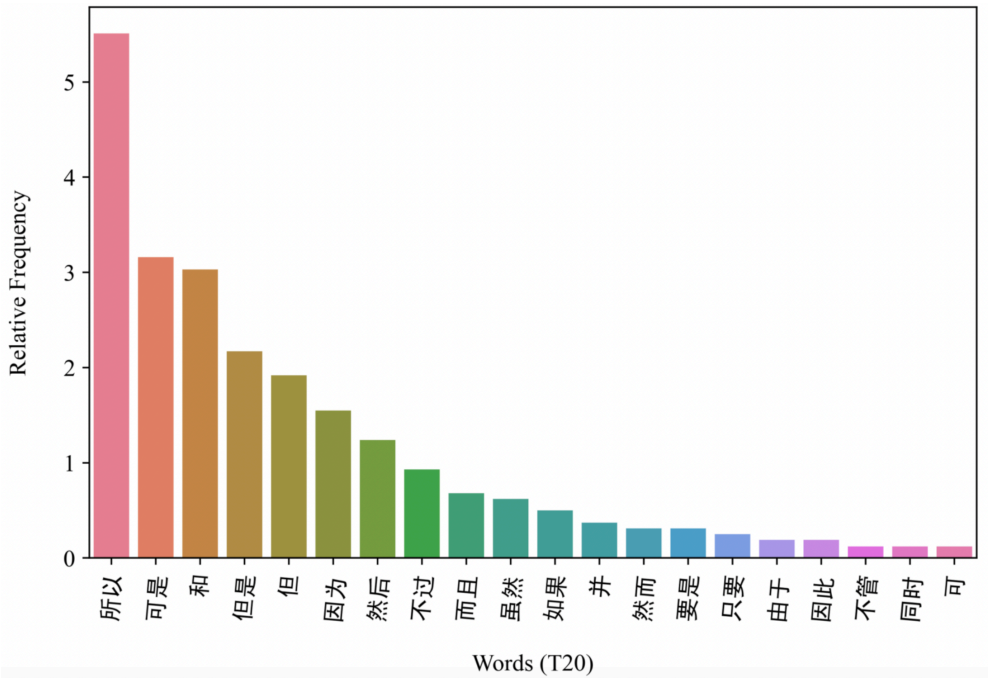

図 4.9　bar plot の拡大図

4.2.3 Wordlist の検索

ワードリスト（Wordlist）検索の手順は上記の品詞（POS）検索手順とほぼ同じである。

（1）左側に示す「Wordlist Search」のチェックボックスにチェックを入れ、「POS Search」の下部にある「Select a wordlist:」のセレクトボックス（select box）から 1 つのワードリストを選択する。ここでは「低难度词汇」（低難度語彙）を例として選択した（図 4.10）。

（2）「JP2」・「JP3」・「CHN」のマルチセレクトボックス（multi-select box、複数選択可）から作文のジャンルを選択する。ここでは「JP2」・「JP3」・「CHN」コーパスにおける“叙事”（出来事の叙述）をジャンルの例として選択した（図 4.10）。

（3）「Wordlist」の下側にある「Word_freq of xxx」のチェックボックスにチェックを入れる。チェックを入れることにより、各ジャンルにおける指定した Wordlist に関する語彙多様性（RTTR）が自動的に表示される（図 4.10）。

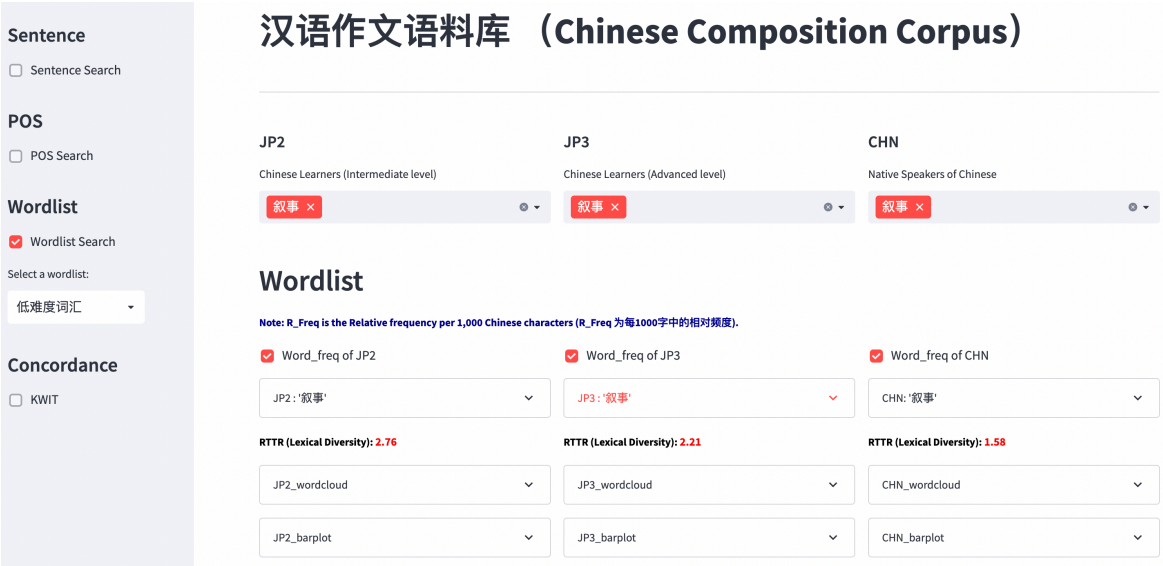

図 4.10　Wordlist の検索画面

（4）「Word_freq of xxx」の下にあるエキスパンダ（expander）領域を押すと、エキスパンダ領域が展開され、指定された Wordlist に属する単語を降順で示す。

例えば、図 4.11 に示すように、3 つのジャンルにおいて、「低难度词汇」に属する単語をエクセルに降順で示している。セルの 1 行目はすべての単語（Tokens）の粗頻度（Freq）と相対頻度（R_Freq）であり、2 行目からは各単語の粗頻度と相対頻度である。「JP2_叙事」・「JP3_叙事」・「CHN_叙事」ジャンルにおける「低难度词汇」の粗頻度（或いは、総頻度・総語数）はそれぞれ、6224 回・13429 回・31609 回であり、1000 字数あたりの相対頻度はそれぞれ、385.51 回・318.12 回・316.90 回である。また、セルの右横にある縦スクロールバーをドロップダウンすると、「JP2_叙事」・「JP3_叙事」・「CHN_叙事」ジャンルにおける「低难度词汇」単語の合計はそれぞれ、218 種・256 種 ・281 種であることが分かった（図 4.12）。

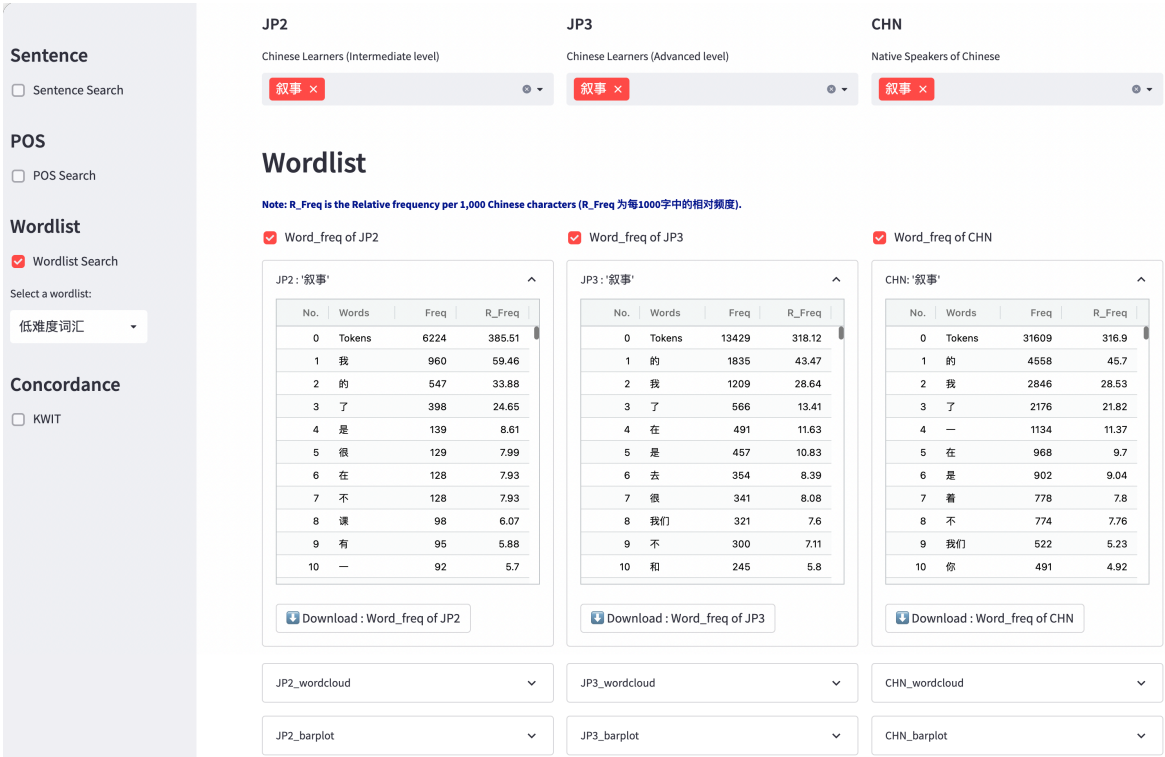

図 4.11　Wordlist：単語の表示

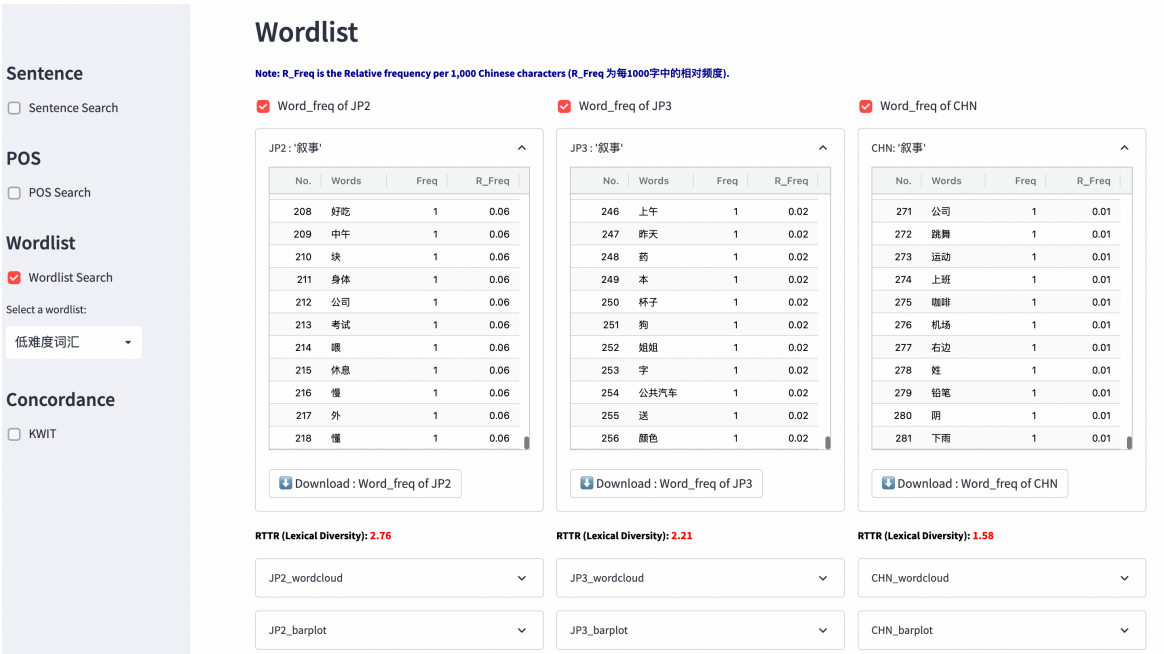

図 4.12　Wordlist：スクロールバーのドロップダウン

図 4.13　Word cloud と barplot の表示

（5）「RTTR (Lexical Diversity): xx」の下にある wordcloud のエキスパンダ（expander）領域を押すと、エキスパンダ領域が展開され、指定した Wordlist に属する単語を wordcloud で示すことができる。また、「wordcloud」の下にある bar plot のエキスパンダ（expander）領域の「＋」を押すと、エキスパンダ領域が展開され、指定した Wordlist に属する単語の上位 20 語を Bar plot で示すことができる（図 4.13）。

4.2.4 Concordance の検索

コンコーダンス（Concordance）検索の機能は、指定したキーワードを含めるテキストや前後の文脈に関する位置情報を特定することができる。コンコーダンスの検索手順は下記の通りである。

（1）左側に示した「Concordance」の下にある「KWIT」のチェックボックスにチェックを入れ、「JP2」・「JP3」・「CHN」のマルチセレクトボックス（multi-select box、複数選択可）から作文のジャンルを選択する。ここでは「JP2」コーパスにおける“叙事”（出来事の叙述）をジャンルの例として選択した（図 4.14）。

(2) 「KWIT (Keyword In Text)」の下側にある text input box にキーワードを入力し、キーワードの左（Left）と右（Right）にある単語数を Slider bar で指定し、「Search」ボタンを押すことにより、キーワードが含まれるセンテンスを自動的に抽出できる（図 4.15）。

図 4.14　Concordance における KWIC の検索

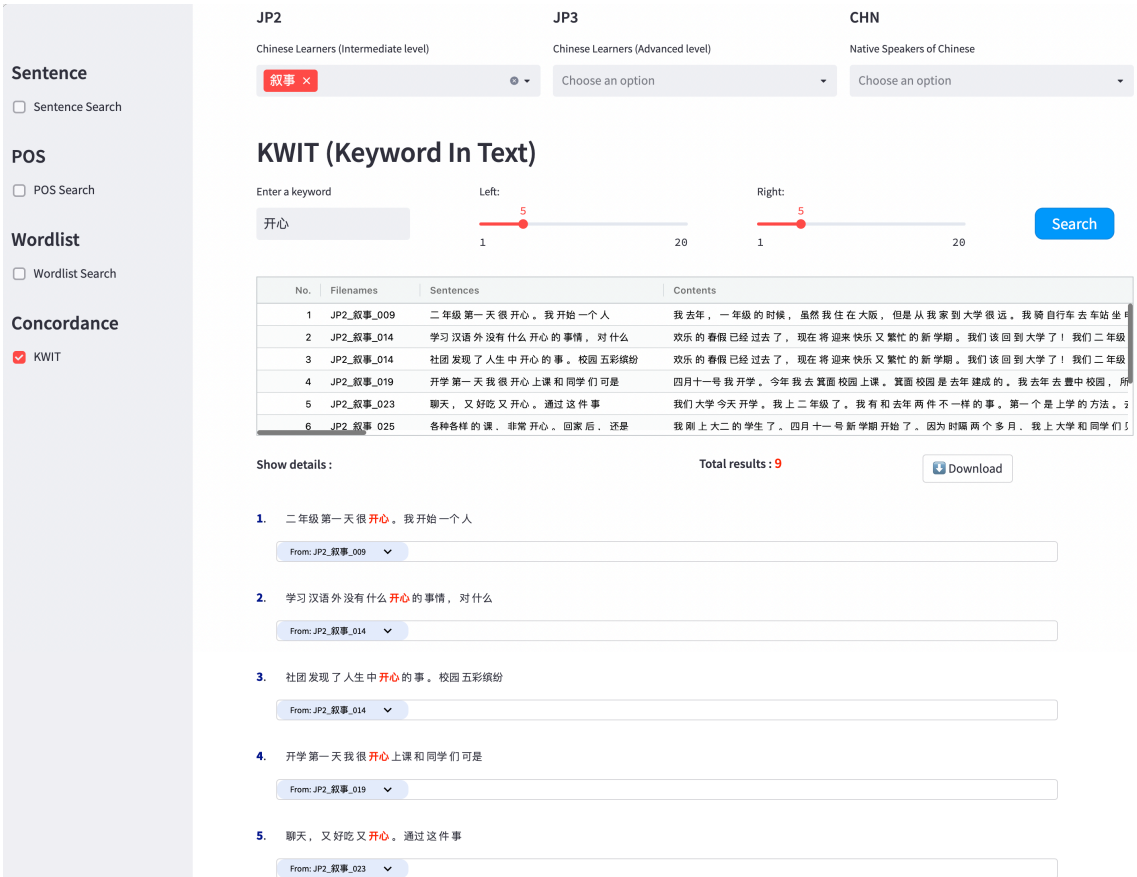

図 4.15　KWIT の検索結果

（3）「Show details」の下にあるエキスパンダ（expander）領域を押すと、エキスパンダ領域が展開され、すべての検索結果は表示できる。センテンスにおけるキーワードの文字色を赤にしており、センテンスの出典も一目瞭然である。例えば、「From: JP2_叙事_009」と「From: JP2_叙事_014」をクリックすると、「JP2_叙事_009」と「JP2_叙事_014」テキストの全体の内容を表示することができ、キーワードを含めるセンテンスが黄色でハイライトされている（図 4.16）。

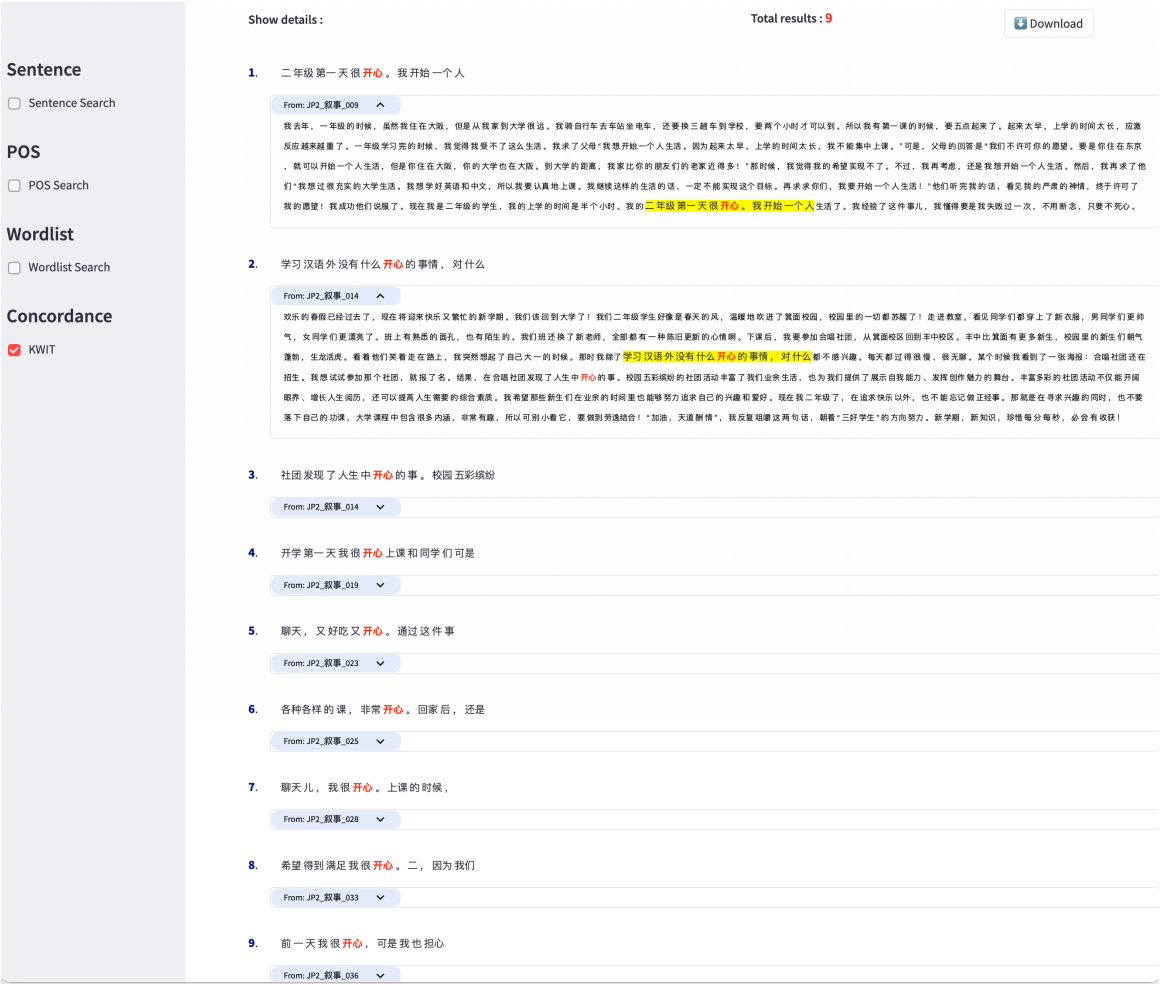

図 4.16　KWIT の検索結果の詳細

（4）左（Left）と右（Right）の Slider bar を変更すると、キーワードの前後の語数を自動的に変更できる（図 4.17）。

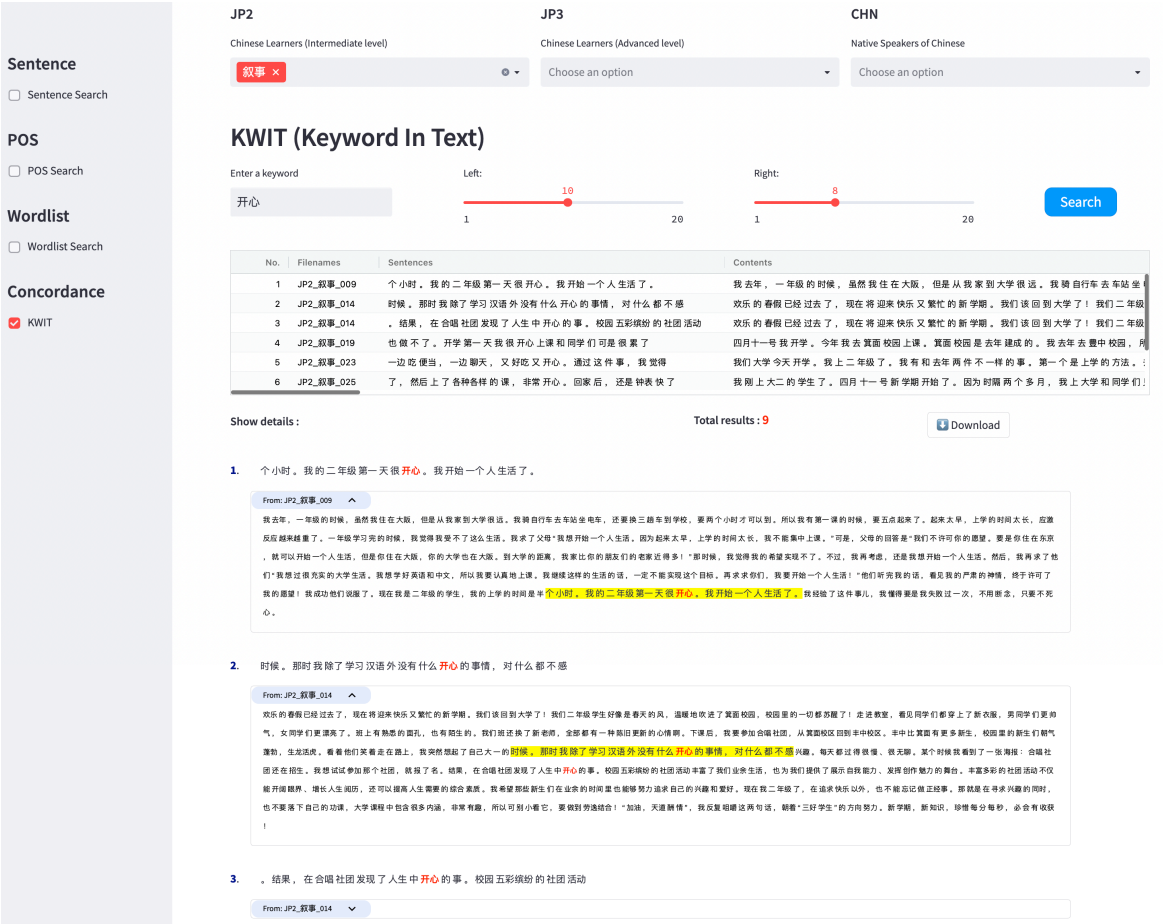

図 4.17　キーワード前後の語数の変更

4.3 本章のまとめ

本章は日本人中国語学習者が書いた作文テキストを用いて構築した学習者の作文コーパスについて説明した。学習者の作文コーパスは、学習者の中国語レベル及び作文のジャンルによってサブコーパス（例えば、JP2_人物描写、JP2_手紙、JP3_人物描写、JP3_意見文）に分類された。また、学習者の作文における言語特徴と比較対照するために、中国語母語話者によって書かれた作文も収集し、構築した母語話者作文コーパスを比較対象とした。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・Wordlist・Concordance などによる検索機能を設けている。これらの検索機能によって、特定のキーワードが含まれるセンテンスの抽出、語彙の使用状況、サブコーパス間の比較などができる。

第 5 章　多次元分析

本章は、多次元分析の手法を用い、中国語母語話者と日本人中国語学習者が書いた中国語作文における言語変異を因子分析で考察する。

5.1 コーパス

多次元分析で使用されるコーパス（データベース）は、第 4 章で構築したコーパスにおける作文テキストである。

5.2 分析手順

データの処理と分析は下記のような手順で行う。

（1）単語分割とタグ付け

各サブコーパスを構成する各テキストに対し、NLPIR[60]で単語分割とタグ付けを行う。単語タグ付けの精度を上げるために、手作業による確認と修正が必要である。下記はタグつけの一例を示している。

去年/t 九月二十二日/t ，/wd 我/rr 去/vf 了/ule 东京/ns 旅行/vi 。/wj 这个/rz 行/ng 的/ude1 最/d 大/a 的/ude1 目的/n 是/vshi 去/vf 看/v 我/rr 非常/d 喜欢/vi 的/ude1 男/b 演员/n 登台/vi 的/ude1 音乐剧/n 。/wj 但是/c ，/wd 上演/v 这个/rz 演剧/vn 的/ude1 剧场/n 在/p 浅草/ns ，/wd 所以/c 我/rr 也/d 期待/v 了/ule 去/vf 这个/rz 地方/n 旅游/vn 。/wj

---

[60] NLPIR は中国語のテキストを分かち書きするツールである。
URL: http://ictclas.nlpir.org/ 〔最終アクセス日：2022 年 11 月 29 日〕

（2）中国語言語項目の選定

多次元分析の段階では、分析用の言語項目群を特定することが重要である。第 3 章で紹介したように、本研究の多次元分析に用いられる言語項目は 10 カテゴリー（例えば、代名詞類、副詞類）に分類される計 92 種の言語項目（例えば、一人称単数代名詞・一人称複数代名詞、否定を表す副詞・程度を表す副詞など）である。

（3）言語項目の頻度集計

Python を用いて各テキストに現れるそれぞれの言語項目の出現頻度を数え、テキスト毎の粗頻度表を作成する。そして、出現頻度を 1,000 字当たりの相対頻度に調整する。第 3 章に示したように、言語項目の粗頻度と相対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のシステムはオンランで一般公開された[61]。

Jiang（2013）は中国語で書かれた文の長さを、単語数より文字数で数えるべきだと述べている[62]。なぜなら、中国語の文章はスペースや単語の区切り記号（句読点を除く）を使わずに書かれるため、中国語の単語の境界が不明確だからであると述べられている（Jiang 2013）。したがって、本研究では、センテンスと単語の長さの測り、及び各言語項目の相対頻度の変換は、単語数ではなく、文字数を基準にする。上記の処理によって得られたデータ表列（1450 行 × 92 列）は、1450 テキストそれぞれに出現する 92 言語項目の標準化頻度（テキスト 1,000 字あたりの相対頻度）である。

（4）データの前処理

多次元分析を行う前、上述の（3）から得られた相対頻度のデータ（1450 行 × 92 列）に対し、相関の高い言語項目及びサブコーパス別における外れ値を除去する必要がある。

ある変数間（本研究では言語項目間）に 0.7 以上の相関がある場合、多重共線性（multicollinearity）を引き起こす可能性がある。92 個の言語項目間の相関係数を求めた結果、「必然性を表す副詞 」と「否定を表す副詞」、「不定代名詞 」と 「疑問代名詞」、「時態助詞“着” 」と「 動作の進行を表す構文」のような 3 つのペア間の相関

---

[61] URL： https://count-linguistic-features.herokuapp.com/

[62] Jiang (2013): One would tend to count *zi* (characters) in any written discourse in Chinese, not only because *zi* is the basic unit in Chinese language but also because the boundary of a Chinese syntactic word is rather unclear, since Chinese is written without using any spaces or other word delimiters (except for punctuation marks).

係数は 0.7 以上であることが分かった。これら 3 つのペアは、強い正の相関である言語項目の同士である。つまり、これら言語項目同士は、お互いに重なり合い、重複する部分が多い。例えば、「動詞＋時態助詞“着”」は「 動作の進行を表す構文」における検索式の一つである。その原因で、「時態助詞“着” 」と「 動作の進行を表す構文」は正の相関が強い言語項目同士になってしまった。相関が高い項目の同士に対して、適当にどちらかの項目を削除する必要がある。本研究では、「必然性を表す副詞 」、「不定代名詞」および「動作の進行を表す構文」の言語項目を削除し、「否定を表す副詞」・「疑問代名詞」・「時態助詞“着”」を保留する。それで、89 種（92−3）の言語項目を説明変数として残っており、この処理で得られたデータは「1450 行 × 89 列」である。

次に、「1450 行 × 89 列」相対頻度のデータに対し、サブコーパス別における外れ値を含んだテキストを除去する。まず、サブコーパス別で各列における相対頻度の数値を Z 得点（Z-score）[63]に変換してみると、5 つの列[64]には、0 で割り算された結果である NaN が生じたため、これら 5 つの列を除去し、「1450 行× 84 列」相対頻度のデータ（下記のコードにおける df_Data）は残された。

そして、下記のコード（図 5.1）で Z 得点絶対値が 5 を超える数値は外れ値として設定され、外れ値を持っている行を自動的に削除することができる。

```
Data = df_Data[df_Data.groupby('Subcorpus').
             apply(lambda x: np.abs(x-x.mean())/x.std() < 5).all(axis=1)]

print(f'Total :  {Data.shape[0]} rows × {Data.shape[1]} columns')
```

```
Total :  1314 rows × 84 columns
```

図 5.1　外れ値を除去するコード及び結果

上記のコードにおいて、「df_Data」は「1450 行× 84 列」の相対頻度データである。上記のコードを実行することで、各サブコーパスにおける外れ値が含まれる行（テキス

[63] Z 得点（Z-score）は標準得点（Standard score）とも呼ばれ、平均（Mean）が 0、標準偏差（Standard deviation, SD）が 1 になるように標準化した値である。

[64] 列名はそれぞれ「他の一人称代名詞」・「感嘆詞」・「擬音語」・「解説複文」・「目的複文」である。

ト）を自動的に除去した。最後は「1314 行 × 84 列」のデータ（上記のコードにおける Data）が得られた。

（5）多次元分析

上記（4）の手順で得られた「1314 行 × 84 列」のデータを多次元分析に用い、Python の Factor_analyzer[65]というパッケージを導入して因子分析を行う。Factor_analyzer において、因子の抽出方法は 3 つがあり、因子の回転手法は 7 種類がある。

① 因子の抽出法：

・minres（ミンレス法）

・maximum likelihood（最尤法）

・principal（主因子法）

Minres（ミンレス法）は最小残差法とも呼ばれ、「所定の数の因子を抽出したあとの残差柑簡を最小にする」（服部 2010）。Maximum likelihood（最尤法）は、「観測変数が多変量正規分布に従うと仮定して因子パターンと独自性を推定する」（服部 2010）。Principal（主因子法）は主成分分析の手法であり、この方法で得られた因子は互いに独立したものではなく、因子モデルの仮定を完全に満たしていない。

② 因子軸の回転方法：

・varimax (orthogonal rotation)：バリマックス法（直交回転）

・promax (oblique rotation)：プロマックス法（斜交回転）

・oblimin (oblique rotation)：オブリミン法（斜交回転）

・oblimax (orthogonal rotation)：オブリマックス法（直交回転）

・quartimin (oblique rotation)：コーティミン法（斜交回転）

・quartimax (orthogonal rotation) ：コーティマックス法（直交回転）

・equamax (orthogonal rotation) ：エカマックス法（直交回転）

---

[65] Factor_analyzer package：https://factor-analyzer.readthedocs.io/en/latest/factor_analyzer.html 〔最終アクセス日：2022 年 10 月 5 日〕

因子軸の回転とは、因子の推定から得られた値を、より良く理解しやすい値に変換する作業だと理解すれば良いと竹内・水本（2012: 167）が述べている。上記の 7 つの回転方法から見られるように、因子軸の回転方法は大きく分けて、斜交回転（oblique rotation）と直交回転（orthogonal rotation）の 2 種類がある（Yong & Pearce 2013）。斜交回転と直交回転の区別について、竹内・水本（2012: 167）は下記ように述べている。

竹内・水本（2012: 167）：

斜交回転（oblique rotation）は、因子間で相関関係が想定される場合に使用する回転である。外国語教育の研究文脈では因子間に何らかの相関関係が見られるのが通常である。また、斜交回転の方が後の因子の解釈が行いやすくなる。斜交回転の代表例としてプロマックス回転（promax rotation）がある。近年は教育心理学などの分野ではプロマックス回転が主流になっている。直交回転（orthogonal rotation）は斜交回転とは逆に因子間の相関をないものとして扱う。代表例としてはバリマックス（varimax rotation）回転がある。

多くの研究は、ほとんどの場合、相関のある因子の仮定がより合理的であると主張している（例えば、Cliff 1987; Fabrigar & Wegener 2012; Pedhazur & Schmelkin 1991; Preacher & MacCallum 2003）。したがって、斜交回転（oblique rotations）が一般に好まれると述べられている（Pituch & Stevens 2016: 345）。

また、Plonsky & Gonulal（2015）が述べているように、人間の認知と言語学習に関わるほとんどの因子は何らかの形で関連していると想定できるため、外国語教育研究において最も適切な選択は一般的に斜交回転（oblique rotation）である。テキストの言語変異に関する研究分野では、因子がテキストの根本的な次元を表しており、因子が完全に無相関であると仮定する理由はないため、斜交回転におけるプロマックス回転（promax rotation）が推奨されると（Biber 1988: 85）は下記のように述べている。

Biber (1988: 85):

There are several different methods of rotation, many of them available as part of standard statistical packages. Two of these are used commonly: Varimax and Promax. These two rotation methods differ in that Varimax maintains orthogonal structure, requiring the assumption that the factors are uncorrelated, while Promax permits oblique structure, that is, it permits minor correlations among the factors. In the description of textual variation, where the factors represent underlying textual dimensions, there is no reason to assume that the factors are completely uncorrelated, and therefore a Promax rotation is recommended.

上で述べた内容理由により、本研究は、minres（ミンレス法）を因子の抽出方法とし、Promax Rotation（プロマックス回転）を因子軸の回転方法として採用した。

（6）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。

5.3 リサーチクエスチョン

本章では、以下のリサーチクエスチョンを立てて考察を行う。

（1）日本人中国語学習者と中国語母語話者が書いた作文において、言語変異を反映する次元があるのか。（第 5 章の 5.4 節と第 6 章の 6.1 節で解明する）

（2）中国語母語話者が書いた作文において、異なる作文ジャンル間に言語変異があるのか。変異があるとすれば、どのような変異があるのか。（第 6 章の 6.2.1 節で解明する）

（3）同一作文ジャンルにおいて、日本人中国語学習者作文と中国語母語話者作文間に言語変異はあるのか。変異があるとすれば、どのような変異があるのか。（第 6 章の 6.2.2 節で解明する）

（4）同一作文ジャンルにおいて、日本人中国語学習者作文と中国語母語話者作文間に言語項目の使用は相違があるのか。相違があるとすれば、どのような変異があるのか。（第 7 章で解明する）

5.4 因子分析

5.4.1 共通性が低い変数の除去

因子分析に使用するデータは「1314 行 × 84 列」（1314 作文テキスト、84 言語項目）である。因子分析[66]を行う前に、KMO と Bartlett の球面性検定（KMO and Bartlett's Test）でサンプルサイズの妥当性を検証する必要がある。

```
# KMO and Bartlett's Test
from factor_analyzer.factor_analyzer import calculate_kmo
kmo_all, kmo_model = calculate_kmo(Data)

from factor_analyzer.factor_analyzer import calculate_bartlett_sphericity
chi_square_value, p_value = calculate_bartlett_sphericity(Data)

print("     KMO :",round(kmo_model, 3), "\n p_value :", p_value,
      "\nchi_square :",round(chi_square_value, 3))
```

```
     KMO : 0.512
 p_value : 0.0
chi_square : 54522.349
```

図 5.2　KMO と Bartlett's Test の結果

KMO と Bartlett の球面性検定を行った結果（図 5.2）、Bartlett の球面性検定の有意確立は $p < 0.001$ であり、KMO 値が 0.512 であることを確認した。得られた KMO 値は標準の 0.6 よりも小さいため（Kaiser 1974 の基準[67]を参照）、共通性（communality）の低い変数（言語項目）を削除する。Biber（2006: 182-183）が述べているように、因子分析を行う前に、一部分の言語項目を削除し整理する必要がある。

> Biber（2006:182-183）：
>
> (1) Some features were dropped because they overlapped to a large extent with other features.
>
> (2) Features were dropped because they were extremely rare.

---

[66] 因子分析用のコードは次のリンクを参考にした。URL：https://medium.com/analytics-vidhya/the-factor-analysis-for-constructing-a-composite-index-2496686fc54c〔最終アクセス日：2022 年 10 月 15 日〕

[67] Kaiser (1974: 35) : 0.90 = marvelous ; 0.80 = meritorious ; 0.70 = middling ; 0.60 = mediocre ; 0.50 = miserable ; below 0.50 = unacceptable.

(3) Some features were dropped because they shared little variance with the overall factorial structure. (‘Communality estimates’ produced by the statistical analysis indicate the extent to which a given features participates in the overall pool of shared variance accounted for by the factor analysis. In general, features with communalities below 0.15 do not have meaningful factor loadings on any factor).

共通性（communality）とは、因子によって説明される各変数の分散の割合である。Biber（2006: 183）は、共通性 0.15 以下の変数がどの因子にも意味のある因子負荷（factor loadings）を持っていないと述べている。言い換えれば、変数（言語項目）の共通性が低ければ低いほどその変数が因子への寄与は低く、共通性が大きいほど因子とは関係の強い変数だといえる。84 種の言語項目における一部分は因子との関係が低いため、Biber（2006: 182-183）が述べている上記の三つの理由に基づいて、一部分の言語項目を削除する。Biber（2006: 183）によると共通性 0.15 未満の変数を削除したほうが良いが、本研究は 0.3 未満の変数（言語項目）を削って因子分析を行う。

図 5.3 に示したコード[68]のように、因子は、ミンレス法（コード：method = ‘minres’）を用いて抽出された。因子がある程度相関していると通常仮定されることを許すという点（Biber 1988: 85）で、因子回転の方法としてプロマックス法（コード：rotation = ‘promax’）が使用された。また、図 5.3 における 3、4 行目のコードを使って、共通性（communality）0.3 以下の言語項目（変数）を自動的に削除し、0.3 より大きい言語項目を自動的に抽出することができる。

```
fa = FactorAnalyzer(rotation='promax',method='minres')
fa.fit(Data)

communalities = pd.DataFrame(fa.get_communalities(), index=list(Data.columns))
features_comm = list(communalities[communalities[0] > 0.30].index)

new_Data = Data[features_comm]
```

図 5.3　共通性（communality）0.3 以下の変数を削除するコード

[68] 「rotation」は軸の回転方法、「rmethod」は因子の抽出方法を設定することである。本研究のコードでは、「rotation」と「rmethod」それぞれを promax （プロマックス回転）と minres （ミンレス法、最小残差法）に設定した。

そして、図 5.4 に示したコードを繰り返して実行し、最後、共通性（communality）が 0.3 より大きい言語項目（変数）は 18 種が得られた。

```
new_Data = Data[features_comm]

fa.fit(new_Data)
communalities = pd.DataFrame(fa.get_communalities(), index=list(new_Data.columns))
features_comm = list(communalities[communalities[0] > 0.30].index)
print('Total variables with communalities > 0.30: {}'.format(len(features_comm)))

# Highlight the values if they are greater than 0.3.
def highlightCommunalities(x):
    return ['background-color: yellow' if v > 0.30 else '' for v in x]

df_communalities = pd.DataFrame(data = {'Variables': new_Data.columns,
                                        'Communality': fa.get_communalities()})
df_communalities.style.apply(highlightCommunalities, subset = ['Communality'])
```

```
Total variables with communalities > 0.30: 18
```

	Variables	Communality
0	高頻度動詞	0.437712
1	動作動詞	0.541921
2	方向動詞	0.384370
3	高頻度副詞	0.941024
4	時間を表す副詞	0.554362
5	程度を表す副詞	0.582677
6	一人称単数代名詞“我”	0.533110
7	時態助詞“着”	0.372860
8	接続詞	0.371718
9	語気詞	0.308464
10	ぼかし言葉	0.568849
11	韻律双音節語	0.326173
12	低難度語彙	0.829626
13	高難度語彙	0.619022
14	非HSK語彙	0.500373
15	語彙多様性	0.607074
16	単語の平均長さ	0.529547
17	平均文長	0.386578

図 5.4　各変数の共通性（communality）を出力するコード及びその結果

5.4.2 KMO と Bartlett の球面性検定

上記 5.4.1 で得られた 1314 行 × 18 列の相対頻度行列（new_Data）を使用して、もう一回 KMO と Bartlett's Test を行った。その結果（図 5.5）、KMO 値は 0.771 であり、Bartlett の球面性検定の有意確立は $p < 0.001$ であったため、1314 行 × 18 列の「new_Data」は、因子分析を行う妥当性があることが確認された。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)と述べられている。本研究の 1314 行 × 18 列の「new_Data」、つまり 1314 テキストと 18 説明変数（言語項目）で構成された行列は、因子分析の条件を満たしていることが分かった。

```
# KMO and Bartlett's Test
from factor_analyzer.factor_analyzer import calculate_kmo
kmo_all, kmo_model = calculate_kmo(new_Data)

from factor_analyzer.factor_analyzer import calculate_bartlett_sphericity
chi_square_value, p_value = calculate_bartlett_sphericity(new_Data)

print("     KMO :", round(kmo_model, 3), "\n p_value :", p_value,
      "\nchi_square :", round(chi_square_value, 3))
```

```
     KMO : 0.771
 p_value : 0.0
chi_square : 13685.163
```

図 5.5　new_Data（1314 行×18 列）の KMO と Bartlett's Test

そして、python における pandas.DataFrame の describe()メソッドで「new_Data.describe()」を実行すると、new_Data（1314 行×18 列）における各数値列の要約統計量（descriptive statistics）を取得できる。表 5.1 は各数値列の要約統計量であり、「new_Data」の各列に対して要素の個数（Count）、平均値（mean）、標準偏差（std）、最小値（min）、1/4 分位数（25%）、中央値（50%）、3/4 分位数（75%）、最大値（max）を示したものである。

表 5.1　new_Data（1314 行 × 18 列）の記述統計量

	Count	mean	std	min	25%	50%	75%	max
高頻度動詞	1314	101.86	20.59	40.00	87.24	100.32	115.16	177.57
動作動詞	1314	56.34	18.01	10.67	43.18	54.76	67.87	133.33
方向動詞	1314	10.68	7.00	0.00	5.65	9.28	15.05	43.96
高頻度副詞	1314	52.09	14.75	9.17	42.33	51.12	61.13	111.64
時間を表す副詞	1314	20.60	9.91	0.00	13.08	19.75	27.29	59.32
程度を表す副詞	1314	12.46	8.21	0.00	6.34	10.65	16.73	50.00
一人称単数代名詞“我”	1314	33.50	21.63	0.00	15.95	33.43	48.02	110.09
時態助詞“着”	1314	3.50	4.43	0.00	0.00	1.98	5.51	24.79
接続詞	1314	21.16	10.00	0.00	13.52	20.16	27.29	59.06
語気詞	1314	9.61	8.11	0.00	3.15	7.77	13.66	48.78
ぼかし言葉	1314	26.01	11.94	0.00	16.91	24.77	33.96	73.83
韻律双音節語	1314	10.24	7.24	0.00	5.00	8.79	14.13	45.98
低難度語彙	1314	331.91	62.31	129.89	294.32	339.67	376.80	501.48
高難度語彙	1314	52.70	23.78	0.00	35.00	50.36	66.22	147.94
非 HSK 語彙	1314	174.85	39.60	66.67	146.20	171.11	197.04	380.28
語彙多様性	1314	10.30	1.89	4.96	8.92	10.13	11.61	16.72
単語の平均長さ	1314	1.51	0.08	1.31	1.45	1.51	1.57	1.85
平均文長	1314	25.62	9.26	8.57	18.77	23.83	30.35	84.29

5.4.3 因子数の決定

続いて、1314 行 × 18 列の相対頻度行列（new_Data）を使用して因子分析を行う。まず、Kaiser-Guttman 基準と Screen Plot で因子数を決定する。Kaiser-Guttman 基準によると、固有値（eigenvalue：因子への影響力の強さ）が 1 以上である因子を選択するべきである。Screen Plot（スクリープロット）は、因子固有値の大きさを視覚的に表現することができ、曲線の折れ曲がりが強い位置までの因子を選択するのが基準である。

因子固有値を出力するコードと各因子固有値の結果は下記の通りである。因子を抽出する際、第 1 因子は共有された分散の最大量を捉え、第 2 因子は第 1 因子が抽出された

後の共有された分散の最大量を捉え、そして以下同様である（Thompson *et al.* 2017）。結果を見やすくために、固有値が 1 より大きい因子は黄色で表示されている。

```
# Check Eigenvalues
EigenValue, value = fa.get_eigenvalues()

# Highlight the values if they are greater than 1.
def highlightEigenvalue(x):
    return ['background-color: yellow' if v > 1 else '' for v in x]

df_eigen = pd.DataFrame({'Factor': range(1, len(EigenValue) + 1), 'Eigenvalue': EigenValue})
df_eigen.style.apply(highlightEigenvalue, subset = ['Eigenvalue'])
```

	Factor	Eigenvalue
0	1	5.016026
1	2	4.147064
2	3	1.535661
3	4	1.087326
4	5	1.017840
5	6	0.766440
6	7	0.647815
7	8	0.547122
8	9	0.519536
9	10	0.441441
10	11	0.426966
11	12	0.368614
12	13	0.356418
13	14	0.332324
14	15	0.307517
15	16	0.258691
16	17	0.172528
17	18	0.050672

図 5.6　因子の固有値を出力するコード及びその結果

図 5.6 は、固有値（eigenvalue）が 1 より大きい因子は 5 つであることを示している。図 5.7 は固有値の大きさをプロットした Screen Plot である。多次元分析の 1 つの手順として、研究者は使用する因子の数を定める必要がある（Thompson *et al.* 2017）。図 5.7 を観察すると、第 1 因子から 5 因子まで、因子の固有値は 1 以上であることを確認した。しかし、推移がなだらかになる前までの因子は 4 番目もしくは 5 番目の因子であると言

える。4 因子と 5 因子の構造を比較した結果、4 因子の構造が最も優れており、抽出された因子の解釈も容易であるため、本研究では、4 因子を採用した。

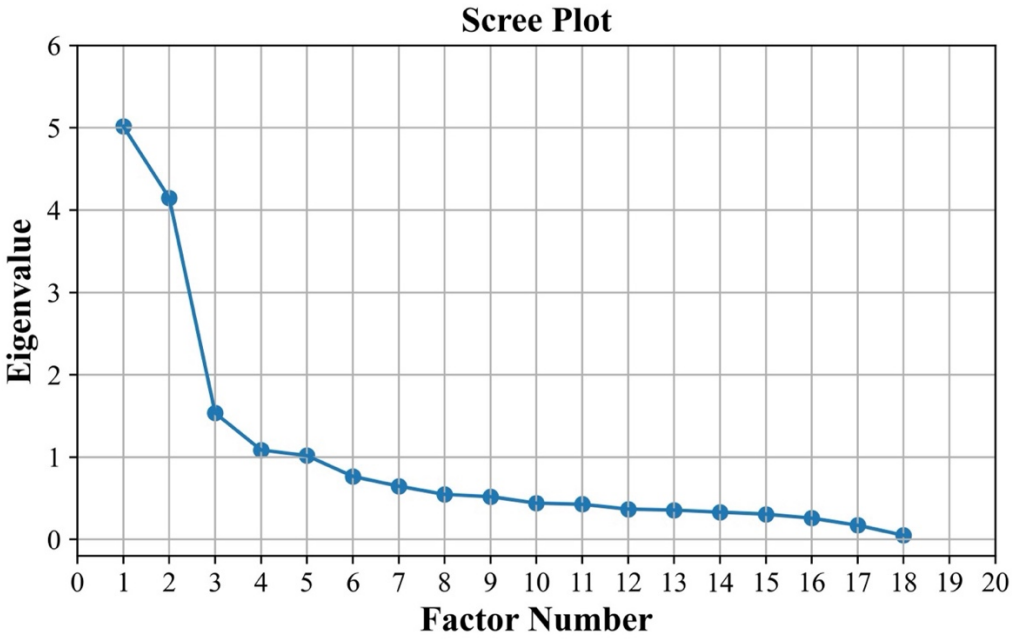

図 5.7　因子の Screen Plot

5.4.4 因子の負荷量

続いて、「fa.loadings_」メソッドで因子負荷量（Factor Loadings）を取得する。図 5.8 は因子負荷量を取得するためのコード及びその結果を示している。4 因子を採用するので、「factor_number = 4」のコードで因子数を設定した。各因子の解釈を容易にするために、プロマックス回転（Promax rataion）を行った。また、「highlightLoadings」の関数を用い、因子負荷量の絶対値が 0.30 より大きい数値は黄色で表示される機能を設けている。下記のコード（図 5.8）を実行したら、因子負荷量を自動的に出力することができた。図 5.8 の結果に示した因子負荷量は、それぞれの言語項目が各因子に与える影響度であり、絶対値が大きいほどその因子への影響が高いことを意味している。

```python
# Factor analysis with 'promax' rotation and 'minres' method.
factor_number = 4
fa = FactorAnalyzer(n_factors = factor_number, rotation = 'promax', method = 'minres')
fa.fit(new_Data)

# Loading factors
fac_loadings = pd.DataFrame(fa.loadings_,
                            columns = ['FAC{}'.format(i) for i in range(1, factor_number+1)],
                            index = new_Data.columns)

# Highlight the values if they are greater than 0.3.
def highlightLoadings(x):
    return ['background-color: yellow' if abs(v) > 0.3 else '' for v in x]

fac_loading_matrix=fac_loadings.style.apply(highlightLoadings)
fac_loading_matrix
```

	FAC1	FAC2	FAC3	FAC4
高頻度動詞	-0.174105	0.013148	0.098985	0.539323
動作動詞	0.131870	0.015472	-0.039234	0.916157
方向動詞	0.130745	-0.091794	-0.049437	0.853418
高頻度副詞	0.154552	-0.547757	1.174826	-0.021973
時間を表す副詞	0.193677	0.171381	0.709916	0.041387
程度を表す副詞	-0.201113	-0.754177	0.157277	0.045434
一人称単数代名詞“我”	-0.853183	0.054474	-0.131128	-0.130865
時態助詞“着”	-0.021426	0.624432	0.029299	0.017067
接続詞	-0.055221	-0.596186	0.035952	-0.053222
語気詞	-0.208098	0.235601	0.180528	0.217695
ぼかし言葉	-0.063486	-0.840374	0.281723	0.061472
韻律双音節語	0.464959	-0.083011	0.026919	-0.174268
低難度語彙	-1.005067	0.065500	0.145831	-0.143110
高難度語彙	0.613895	0.277319	0.016917	-0.165214
非HSK語彙	0.556531	0.344564	-0.079418	0.182605
語彙多様性	0.580975	0.389995	0.126566	-0.004576
単語の平均長さ	0.434289	-0.518553	-0.347096	0.017766
平均文長	0.479686	0.225569	0.222237	-0.102111

図 5.8　因子負荷量を出力するコード及びその結果

5.4.5 因子間の相関

4 つの因子間の Pearson 相関係数を求めるコードとその結果は図 5.9 に示した。図 5.10 は因子間相関の強さを Heatmap で可視化したものである。Heatmap は、行列の個々の値を色や濃淡で可視化することができる。図 5.9 と図 5.10 に示すように、最大の因子間相関は、因子 2 と 3（-0.179）、因子 3 と 4（-0.168）、因子 1 と 2（0.152）である。因子間の相互相関はとても小さかったため、4 つの因子はお互いに独立している因子だと言える。

```
# Inter-factor correlation
Corr_fac_loadings=fac_loadings.corr(method ='pearson')
Corr_fac_loadings
```

	FAC1	FAC2	FAC3	FAC4
FAC1	1.000000	0.151797	0.003450	0.032772
FAC2	0.151797	1.000000	-0.179306	0.029413
FAC3	0.003450	-0.179306	1.000000	-0.168110
FAC4	0.032772	0.029413	-0.168110	1.000000

図 5.9　因子間の相関を出力するコード及びその結果

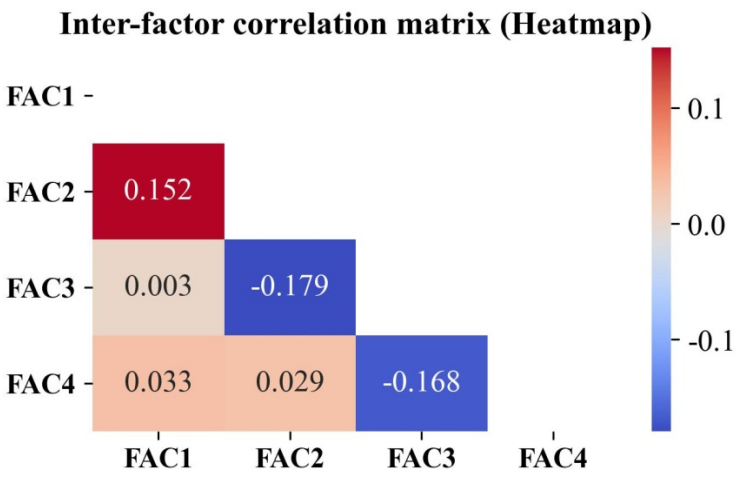

図 5.10　因子間相関の Heatmap 図

5.4.6 各因子の寄与率

最後に、「fa.get_factor_variance()」の機能で、各因子の寄与度（SS Loadings）、寄与率（Proportion Variance）、累積寄与率（Cumulative Variance）を出力することができる。

```
# Explained variance
idx = ['SS Loadings', 'Proportion Variance', 'Cumulative Variance']
df_variance = pd.DataFrame(data = fa.get_factor_variance(),
                           index = idx,
                           columns = ['FAC{}'.format(i)
                                      for i in range(1, factor_number+1)])
df_variance
```

	FAC1	FAC2	FAC3	FAC4
SS Loadings	3.614584	3.095772	2.268486	2.056467
Proportion Variance	0.200810	0.171987	0.126027	0.114248
Cumulative Variance	0.200810	0.372798	0.498824	0.613073

図 5.11　因子の累積寄与率

各因子の寄与率（Proportion Variance）については、第 1 因子は 20.08%、第 2 因子は 17.20%、第 3 因子は 12.60%、第 4 因子は 11.42%であり、4 つの因子の累積寄与率（Cumulative Variance：第 1 因子から第 4 因子まで順番に加算した寄与率）は約 61.31%である。言い換えれば、4 つの因子でデータ全体の約 61.31%を解釈することができる。Field（2009）は、モデル化された総分散の累積パーセンテージ（cumulative percentage）が、最低でも 55%～65%程度であるべきであることを提案している。本研究のデータでは、抽出された因子は全分散の 61.33%を説明でき、Field の提案に合致していることが明らかになった。また、この値は、Biber（1988）による 7 因子解の 51.9%、Friginal & Weigle（2014）による 4 因子解の 30.5%、Egbert & Biber（2016）による 10 因子解の 42.7%、のような他の多次元分析の研究よりも、累積寄与率が高いことが確認された。

5.4.7 共起パターン

図 5.8 に示された因子負荷量（Factor Loadings）を小数点以下 2 桁まで四捨五入して、図 5.12 の Heatmap（ヒートマップ）で可視化する。

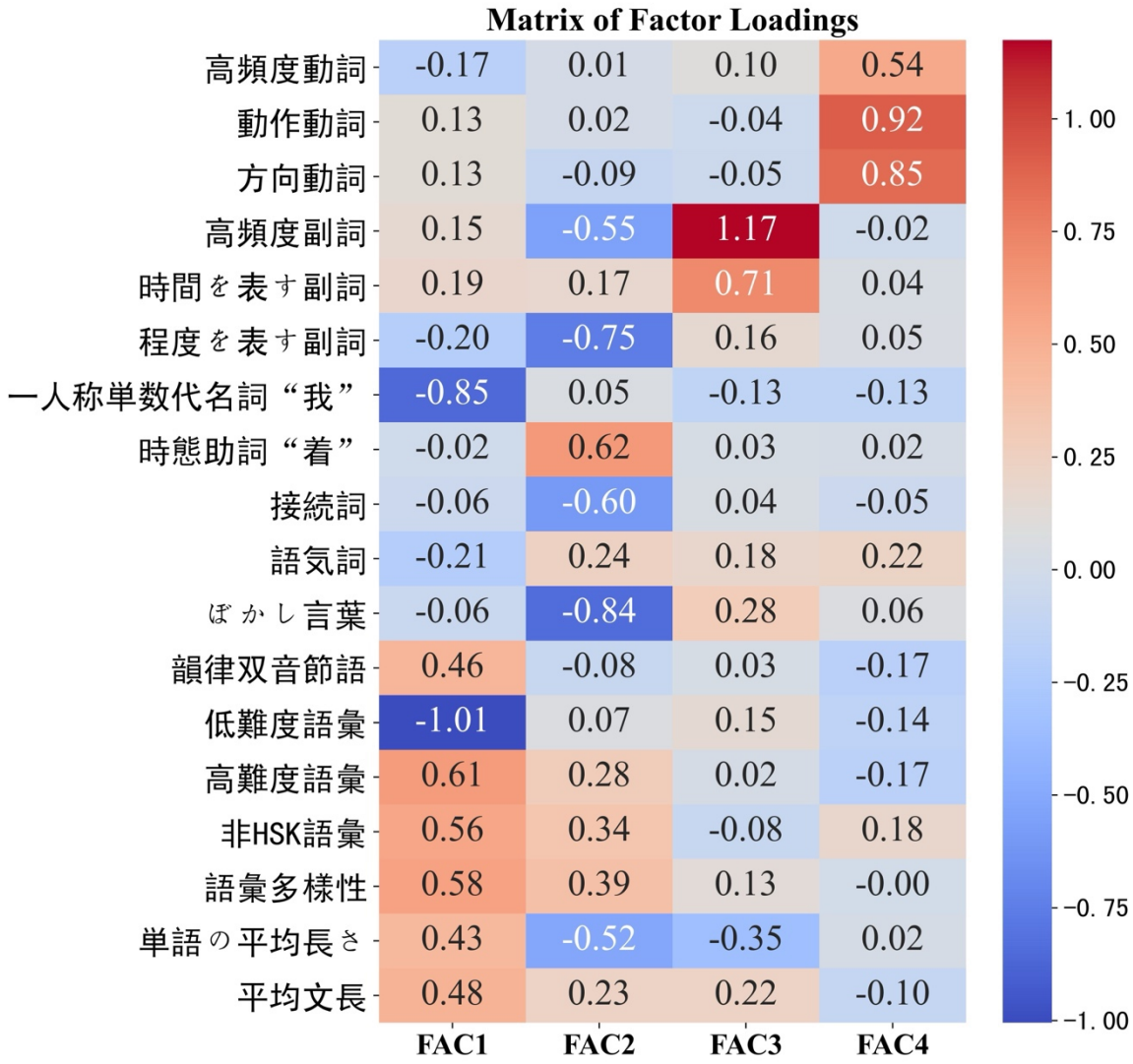

図 5.12　因子負荷量の Heatmap 図

Heatmap は、前述のように、行列の個々の値を色や濃淡で可視化することができる。図 5.12 に示したように、各因子列における言語項目負荷量の絶対値は低ければ低いほど色が淡くなり、なお、その言語項目はその因子への寄与が低い。それに対して、各因子列における言語項目の負荷量の絶対値は高いほど色が濃くなり、その言語項目はその因子への寄与率が高いことを意味している。例えば、「FAC1」において、因子の正次元への寄与が高い言語項目は「高難度語彙」・「語彙多様性」・「非 HSK 語彙」・「平均文長」・「韻律双音節語」などであり、負次元への寄与が高い言語項目は「低難度語彙」・「一人称単数代名詞“我”」などである。

表 5.2 は図 5.12 によって整理した言語項目間の共起パターン（Co-occurrence patterns）である。上記の因子負荷量は因子軸を Promax Rotation で回転した後の Pattern Matrix の結果であるため、負荷量の大きさが 1 を超えることが可能である[69]。また、前述のように、言語項目に与えられた因子負荷量の絶対値が大きければ大きいほどその因子への貢献が大きい。ただし、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（Biber 1988: 87）[70]。そのため、表 5.2 には、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。

その他、いくつかの言語項目は、複数の因子に顕著な負荷量（絶対値の大きさが 0.30 以上）を持っている。例えば、「単語の平均長さ」は Factor 1・Factor 2・Factor 3 に重複して現れており、「高頻度副詞」は Factor 2・Factor 3 に重複して現れている。複数の因子に重複で現れた言語項目について、Gorsuch（1983: 268）と Biber（1988: 93）は以下のように述べている。

> Gorsuch（1983: 268）：
>
> For each variable, examine all of its salient loadings on all factors to find that factor with which the variable has its strongest relationship. The variable is then used to measure only that factor; it is treated as a non-salient variable on all other factors.
>
> Biber（1988: 93）：
>
> [E]ach linguistic feature is included in the factor score of the factor on which it has the highest loading (in terms of absolute magnitude, ignoring plus or minus sign).

---

[69] 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（リンク： In Factor Analysis (or in PCA), what does it mean a factor loading greater than 1? - Cross Validated (stackexchange.com); 最終アクセス日：2022 年 10 月 20 日）。また、Watkins（2018: 233）は「pattern coefficients might exceed |1.00|」と述べている。

[70] Biber（1988: 87）: [L]oadings having an absolute value less than .30 are generally excluded as unimportant even if they are statistically significant. Only the important, or “salient”, loadings should be interpreted as part of each factor.

表 5.2　言語項目間の共起パターン

Factor 1		
高難度語彙	0.61	
語彙多様性	0.58	
非 HSK 語彙	0.56	
平均文長	0.48	
韻律双音節語	0.46	
単語の平均長さ ）	0.43	）
……		
一人称単数代名詞“我”	-0.85	
低難度語彙	-1.01	
Factor 2		
時態助詞“着”	0.62	
語彙多様性 ）	0.39	）
非 HSK 語彙 ）	0.34	）
……		
単語の平均長さ	-0.52	
高頻度副詞 ）	-0.55	）
接続詞	-0.60	
程度を表す副詞	-0.75	
ぼかし言葉	-0.84	
Factor 3		
高頻度副詞	1.17	
時間を表す副詞	0.71	
……		
単語の平均長さ ）	-0.35	）
Factor 4		
動作動詞	0.92	
方向動詞	0.85	
高頻度動詞	0.54	

表 5.3 では、因子得点の計算に使われない言語項目が括弧内に示されている。つまり、括弧が付けられている言語項目は、「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勘定されない」（田畑 2002: 83）。

例えば、「単語の平均長さ」は Factor 1 への負荷量が「0.43」、Factor 2 への負荷量が「-0.52」、Factor 3 への負荷量が「-0.35」であるため、Factor 1 と Factor 3 より Factor 2 の方が負荷量の絶対値が高い。従って、因子得点の計算や因子機能を解釈する際、「単語の平均長さ」は Factor 1 と Factor 3 に使用されず、Factor 2 に使用する。同様に、「高頻度副詞」は Factor 2 への負荷量が「-0.55」、Factor 3 への負荷量が「1.17」であるため、「高頻度副詞」は Factor 2 に使用されず、Factor 3 の因子得点の計算や因子機能の解釈に使われる。

5.4.8 次元スコア

次に、各テキストの次元スコアを算出する。次元スコア（Dimension score）を計算する前に、各列における相対頻度を平均値が「0」、標準偏差（standard deviation）が「1」となるような Z-score （Z 得点）に標準化する必要がある。これは、個々の頻度に関係なく、すべての言語項目がテキストの次元スコアに等しい影響を与えることを保証するために行われた処理である。Z-score の計算方法は、該当する相対頻度から平均値（mean）を引き、標準偏差（standard deviation）で割ることである。Z-score の計算公式は下記の通りである。

$$z = \frac{x - \mu}{\sigma}$$

上記の公式において、x はある言語項目の相対頻度、μ はその項目のコーパス全体での平均値、σ はその項目のコーパス全体での標準偏差である。テキストの言語項目を次元ごとに標準化した後、これら標準化数値を加算して得られた値を次元スコアと呼ぶことにする。つまり、本研究のコーパスを構成する各テキストには 4 つの次元のスコアが付与されている。次の節では、次元スコアを python で自動的に計算する方法を紹介する。

5.4.8.1 各テキストの次元スコア

各言語項目の平均値と標準偏差を表 5.1 に示している。相対頻度が Z-score に標準化されると、新しい尺度に変換される。表 5.3 には、Factor 1 の次元スコアの計算に使用され

る 7 つの言語項目は表示している。なお、表 5.2 における括弧内の項目「単語の平均長さ 」は Factor 2 の計算に使用されるべきである。Z-score の算出について、作文テキスト「JP3_人物描写_008」と「CHN_人物描写_037」を例として説明する。例えば、「非 HSK 語彙」を考えてみよう。表 5.1 から、この言語項目の平均値は 174.85、標準偏差は 39.60 であることが分かった。また、テキスト「JP3_人物描写_008」と「CHN_人物描写_037」において「非 HSK 語彙」の出現頻度が 1,000 字あたり 164.67 回と 187.01 回であったため（表 5.3）、この 2 つのテキストの「非 HSK 語彙」に関する Z-score は下記のように求められる。

「JP3_人物描写_008」：

（164.67－174.85）/ 39.60 = －0.26

「CHN_人物描写_037」：

（187.01－174.85）/ 39.60 = 0.31

表 5.3　例文の相対頻度と Z-score

	File_name	高難度語彙	語彙多様性	非 HSK 語彙	平均文長	韻律双音節語	一人称単数代名詞“我”	低難度語彙
相対頻度	JP3_人物描写_008	22.46	9.38	164.67	20.88	5.99	37.43	342.81
	CHN_人物描写_037	21.65	10.34	187.01	29.88	7.87	0.00	338.58
Z-score	JP3_人物描写_008	-1.27	-0.49	-0.26	-0.51	-0.59	0.18	0.18
	CHN_人物描写_037	-1.31	0.02	0.31	0.46	-0.33	-1.55	0.11

上記のように各テキストに対し、因子を構成する各言語項目の出現頻度の Z-score を求める。各テキストの次元スコアの算出方法は下記の通りである。

各テキストの次元スコア ＝「正の言語項目の Z-score 合計」－「負の言語項目の Z-score 合計」

つまり、次元スコアは、正の負荷がかかる言語項目の標準化スコア（Z-score）をすべて足し、負の負荷がかかる言語項目の標準化スコアをすべて引くことによって計算される。例えば、Factor 1 に属する言語項目において、正の因子負荷量を持っている言語項目は「非 HSK 語彙」・「語彙多様性」・「高難度語彙」・「平均文長」であり、負の因子負荷量を持っている言語項目は「一人称単数代名詞“我”」・「低難度語彙」であるため、テキスト「JP3_人物描写_008」と「CHN_人物描写_037」の次元スコアは次のように求められる。「JP3_人物描写_008」の 次元 1 スコアは − 3.48 であり、「CHN_人物描写_037」の 次元 1 スコアは 0.59 である。

「JP3_人物描写_008」：

$$(-1.27) + (-0.49) + (-0.26) + (-0.51) + (-0.59) - (0.18 + 0.18) = -3.48$$

「CHN_人物描写_037」：

$$(-1.31) + 0.02 + 0.31 + 0.46 + (-0.33) - (-1.55 + 0.11) = 0.59$$

Factor 1〜4 において、各テキストの次元スコアは、python によって自動算出できる。下記の図 5.13 には 、各テキストに対し、Factor 1 での次元スコアを出力するコードを例として示している。図 5.13 のコードにおいて、「new_Data」は因子分析に使用された 1314 行×18 列の相対頻度行列である。次元スコア計算のステップとしては、まず「new_Data」の各列に対して Z-score を求め、そして因子を構成する言語項目を負荷量の「正」と「負」グループに分け、最後に「正の負荷量を持っている言語項目の Z-score 合計」から「負の負荷量を持っている言語項目の Z-score 合計」を引いて次元スコアを算出することである。

```
# Calculate Z-scores
new_Data_Zscore = new_Data.apply(lambda x:(x-x.mean())/x.std())

# Compute factor scores of Factor 1
Fac1_positive = ['高難度語彙','語彙多様性','非HSK語彙','平均文長','韻律双音節語']
Fac1_negative = ['一人称単数代名詞“我”','低難度語彙']

Fac1_scores = pd.DataFrame(new_Data_Zscore.apply(
    (lambda x : x[Fac1_positive].sum() - x[Fac1_negative].sum()), axis = 1),columns = ['Factor1_scores'])
Fac1_scores
```

図 5.13　例：次元スコアを算出するコード

5.4.8.2 各ジャンルの次元スコア

Factor 1 から Factor 4 まで、各テキストの次元スコアを算出した後、図 5.1 に示したようなコードを用い、サブコーパス（Subcorpus）別で次元スコアの外れ値（Z-score の絶対値 > 5）が含まれている行を除去する。因子分析に使用されたデータは 1314 行であったが、次元スコアの外れ値を除去する操作によって、最後は 1312 行のデータが残された。

中国語母語話者が書いた作文における各ジャンル（サブコーパス）間の関係と、中国語学習者・中国語母語話者書いた作文におけるジャンル間の比較は、各ジャンルの次元スコアの平均値を求めることで考察することができる。各次元において高い正のスコアを持つテキストまたはジャンルは、その次元に関連する正の言語項目を多く出現し、負の言語項目をほとんど出現していない。一方、正の次元スコアが低い場合、テキストまたはジャンルが、その次元の負の項目を多く含んで、正の項目をほとんど含まないことを意味している。

表 5.4 は、1312 行のデータに対し、次元 1 から次元 4 までの次元スコアをサブコーパス別（Subcorpus）の平均値で示している。

表 5.4　次元スコアの平均値（Mean Scores of Dimensions）

Corpus	Subcorpus	N	次元 1	次元 2	次元 3	次元 4
中国語母語話者	CHN_人物描写	137	0.33	3.91	1.26	2.08
	CHN_出来事の叙述	133	2.37	3.49	0.80	1.38
	CHN_意見文	132	8.03	1.15	-0.46	-1.63
	CHN_手紙	136	2.00	2.27	0.55	-0.42
	CHN_日記	136	2.66	2.22	0.95	0.35
日本人中国語学習者	JP2_人物描写	163	-5.01	-2.63	-0.96	-0.86
	JP2_出来事の叙述	40	-6.00	-1.17	-0.74	0.89
	JP2_手紙	74	-4.20	-3.15	-0.22	-0.45
	JP2_日記	142	-3.30	-3.11	-0.80	0.40
	JP3_人物描写	103	-3.26	-2.00	0.02	-0.90
	JP3_出来事の叙述	59	-0.29	-3.07	-1.07	-0.12
	JP3_意見文	57	2.11	-4.05	-0.85	-1.06

5.5 本章のまとめ

本章では、因子分析の技術的概要を紹介した上で、多次元分析のための因子抽出、因子負荷量の出力と次元スコアの計算を行った。次章では、本章で得られた 4 つの因子（言語項目の共起パターン）を主な次元として捉え、各因子を構成する有効な言語項目の広く共通な伝達機能を吟味しながら、各次元を機能的に解釈し、サブコーパス間の類似点と相違点を分析する。

第 6 章　中国語作文における言語変異の次元

第 5 章では、因子分析によってテキスト間に共起する言語項目群を 4 つの因子として特定した。表 5.2 の共起パターンは、便宜上、表 6.1 として下記に繰り返して表示される。因子の機能を解釈するために、本章では、4 つの因子を言語変異の主な次元（Dimension）として捉える。

表 6.1　言語項目の共起パターン

Dimension 1	
高難度語彙	0.61
語彙多様性	0.58
非 HSK 語彙	0.56
平均文長	0.48
韻律双音節語	0.46
単語の平均長さ ）	0.43 ）
……	
一人称単数代名詞“我”	-0.85
低難度語彙	-1.01
Dimension 2	
時態助詞“着”	0.62
語彙多様性 ）	0.39 ）
非 HSK 語彙 ）	0.34 ）
……	
単語の平均長さ	-0.52
高頻度副詞 ）	-0.55 ）
接続詞	-0.60
程度を表す副詞	-0.75
ぼかし言葉	-0.84
Dimension 3	
高頻度副詞	1.17
時間を表す副詞	0.71
……	
単語の平均長さ ）	-0.35 ）
Dimension 4	
動作動詞	0.92
方向動詞	0.86
高頻度動詞	0.54

6.1 次元の解釈

言語使用域（registers）は、テキストに共通して見られる言語項目の組み合わせからなる変異の次元（統計的因子分析によって特定された共起パターン）に関して比較される。「各次元は 3 つの点で特徴的である」と Biber（2014）[71]は下記のように説明している。

1）共起する言語項目の明確なセットによって定義される。
2）特定の伝達機能に関連する。
3）それぞれの次元に関連する言語使用域の変異には、異なるパターンがある。

表 6.1 に示したように、4 つの次元の大半は、主に因子負荷量の正（Positive loadings）と負（Negative loadings）で 2 つのグループに分けられている。ここで強調したいのは、「正：Positive」と「負：Negative」はこの分析における評価用語ではなく、むしろ連続体の両極を表していることである（Friginal & Weigle 2014）。「正」と「負」のグループについて、Biber（1988: 101）と Biber *et al*（2002）は下記のように述べている。

> [T]he features with positive weights co-occur in texts; the features with negative weights co-occur in texts; and these two groups of features occur in a largely complementary distribution. That is, when a text has several occurrences of the features with negative weights it will likely have few of the features with positive weights, and vice versa.
>
> Biber（1988: 101）

> Features with higher loadings are thus better representatives of the dimension underlying a factor. Most of the dimensions consist of two groupings of features, one with positive and the other with negative loadings. The positive and negative sets represent features that occur in a complementary pattern. That is, when the features in one group occur together frequently in a text, the features in the other group are markedly less frequent in that text, and vice versa.
>
> Biber *et al*（2002: 14）

[71] Each dimension is distinctive in three respects: (1) It is defined by a distinct set of co-occurring linguistic features; (2) It is associated with particular communicative functions; (3) There are different patterns of register variation associated with each dimension (Biber 2014).

つまり、「正」と「負」グループ別に属する言語項目はテキスト内に共起する傾向が強い。「正」と「負」のグループは、相補的に出現する言語項目同士の組み合わせである。「あるテキストで、一方のグループに属する項目の共起する頻度が高いときには、もう一方のグループに属する項目は、そのテキストで明らかに共起する頻度が低い。逆もまた同様である」（斉藤ら訳 2003: 161）と述べられている。

まず、「正」のグループにおける言語項目群はテキストでの共起傾向が強い。例えば、次元 1 において、「正」のグループ[72]を構成した言語項目は「高難度語彙」・「語彙多様性」・「非 HSK 語彙」・「平均文長」・「韻律双音節語」であるため、あるテキストには「高難度語彙」の出現回数が多ければ、「非 HSK 語彙」や「韻律双音節語」の出現回数も多く、「語彙多様性」が高く、「平均文長」が長い可能性も高い。

同様に、「負」のグループに属する言語項目群はテキスト内に共起する傾向が強い。次元 1 の負のグループに属する言語項目は「低難度語彙」・「一人称単数代名詞“我”」であるため、あるテキストには、「低難度語彙」が頻繁に現れる時、「一人称単数代名詞“我”」も多く現れる可能性が高い。

また、「正」と「負」のグループにおける言語項目は「相補的分布をなす」と斉藤ら訳（2003: 162）が述べているように、あるテキストには、正の言語項目（例、「高難度語彙」、「非 HSK 語彙」）が頻繁に現れる場合、負の言語項目（例、「低難度語彙」、「一人称単数代名詞“我”」）の出現頻度は明らかに少ない。それに対し、あるテキストには正の言語項目の出現頻度が低ければ、負の言語項目は明らかに頻繁に現れるのが一般的である。

因子分析で特定した共起パターンは、各言語項目が広く共通する伝達機能（communicative function）の根底にあるものを示している。言い換えれば、言語項目がテキストに共起するのは、それらの項目が共通の伝達機能を果たすためにテキストで使用されるからである。次元を機能的に解釈する基礎については、斉藤ら訳（2003: 161）は、「次元を規定している共起項目の組み合わせに広く共通する機能を分析すること」、「次元という観点からレジスター間の類似点と相違点を分析すること」、「次元を解釈する第 1 歩は、共起する項目に共通の機能を見つけ出すことである」と述べている。

---

[72] 「合偶双音節語」と「単語の平均長さ」それぞれは次元 1 と次元 4 に重複して現れており、次元 1 と次元 4 における負荷量の方が高い。

例えば、次元 1 では、「正」と「負」グループ別に属する言語項目の特徴を考えると、比較的解釈しやすい。ある文章に「低難度語彙」と「一人称単数代名詞“我”」は頻繁に現れると、この文章は主に低難度語彙に限定して一人称“我”（私）の視点で書かれた文であることがすぐに分かるだろう。それに対して、ある文章は語彙多様性が高く、センテンスが長く（平均文長）、また「高難度語彙」・「非 HSK 語彙」が頻繁に出現する場合、この文章は豊富な語彙で書かれたものであると考えられる。下記の例文(1)～（3）に示した学習者と中国語母語話者が書いた作文の抜粋において、次元 1 の負のグループに属する 2 つの言語項目が共起することを示している（注：低難度語彙には下線を引いてある）。

---

(1) 我 想 马上 前往 新 箕面校园 ，不过 想 起来 我 不 知道 怎么 去 那里 。 立刻 我 打算 用 手机 查 怎么 去 的 时候 ， 发觉 我 忘 了 带 手机 。 我 查 不 了 怎么 去 。 我 不 想 问 站务员 、 行者 。 那 怎么办 哪？

（JP2_出来事の叙述_001, Dimension 1 score: -6.81）

(2) 第二天 的 二十二日 的 早晨 ， 我 到 了 东京 车站 。 我 去 了 麦当劳 ， 吃了 早饭 。 休息 了 一会儿 之后 ， 我 坐 电车 浅草 。 演剧 下午 一点 开演 ， 所以 我 有 一点儿 时间。浅草 是 日本 有名 的 旅游 景点 ， 有 很多 从 外国 来 的 观光客 。 我 先 去 了 浅草寺 ， 这 以 “ 雷门 ” 而 有名 。 我 在 这儿 抽 了 签 ， 结果 是 吉 。 我 也 仍 了 香钱 ， 祈祷 了 我 的 大学 的 成绩 提高 。 然后 ， 我 散步 浅草寺 的 附近 的 商品街 。 商店卖 小吃 、 土产 等等 。 我 买 了 “ 炸糕 ” ， 浅草 的 名产 。 炸糕 非常 好吃 ， 所以 我 竟 吃 了 三 个 。 我 也 去 了 卖 土产 的 商店 ， 买了 给 家族 的 脆饼 。

（JP3_出来事の叙述_052, Dimension 1 score: -4.42）

(3) 阳光 透过 窗子 也 照 进 教室 里 ， —— 叮 叮 叮 ！ 上课 了 。 正在 外面 做 游戏 的 同学 们 ， 听到 上课 铃 立马 走 进 教室 回到 座位 上 ， 可 教室 里 仍 没有 完全 平静 ， 直到 老师 走 到 讲台 上 拍 着 桌子 ， 说 ： “ 安静 ！ 开始 念 成绩 ！ ” 老师 一 开始 心情 非常 好 ， 可是 念 到 后面 时 ， 突然

皱 起 眉头 ， 眼睛 瞪 了 一下 我 身后 的 他 ， 他 正在 摇晃 桌子 ， 老师 摇 摇 头 ， 轻 叹 口 气 念 出 了 他 的 成绩 。 这个 我 最 讨厌 的 他 竟然 考 了 全班 倒数 第三 名 ， 我 坐 在 座位 上 ， 美滋滋 地 笑 了 。

（CHN_出来事の叙述_007, Dimension 1 score: 2.77）

上記の 3 つの例文は典型的な出来事の叙述（“叙事”）の例で、一人称の視点からある場所で経験したことに関する叙述である。全体的に見ると、(1)～（2）の例文は、センテンスの長さが短く、一人称代名詞“我”のような難度の低い語彙を頻繁に使用している。例文(1)～（2）に比べ、例文（3）は長いセンテンスで書かれ、一人称代名詞“我”のような難度の低い語彙が多く現れていないことが分かった。

表 6.1 に示した 4 つの主な次元は、中国語母語話者と学習者が書いた作文のジャンル（表 4.1 のサブコーパス）間の類似点や相違点を表している。各次元に共起する言語特徴の機能に基づき、4 つの次元を以下のようにラベリングしている。

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

下記では、5.4.8.2 節から求められた各テキストの次元スコア（表 5.4）に基づき、共起する言語項目群の広く共通な伝達機能を分析するとともに、サブコーパスに見られる言語変異の相を分析していく。

6.1.1 次元 1：産出語彙の豊富さ VS. 産出語彙の限定性

次元 1 に属する「単語の平均長さ」は次元 1 ・次元 2・次元 3 に重複して現れており、次元 2 における負荷量の方が高いため、この言語項目は次元 1 の解釈に使用されない。すなわち、次元 1 をラベリングする際、主に正の負荷量を持っている言語項目「高難度語彙」、「語彙多様性」、「非 HSK 語彙」、「平均文長」、「韻律双音節語」と負の

負荷量を持っている言語項目「一人称単数代名詞“我”」・「低難度語彙」で解釈を行う。表 6.2 は次元 1 の解釈に使用される言語項目しか表示していない。

表 6.2　次元 1 に属する言語項目

Dimension 1	
高難度語彙	0.61
語彙多様性	0.58
非 HSK 語彙	0.56
平均文長	0.48
韻律双音節語	0.46
……	
一人称単数代名詞“我”	-0.85
低難度語彙	-1.01

図 6.1 は、次元 1 において、各サブコーパスの次元スコアの平均値を昇順に示した横棒グラフの可視化図である。図 6.1 のような横棒グラフにより、サブコーパス間の比較や次元の解釈をもっとわかりやすく説明することができる。

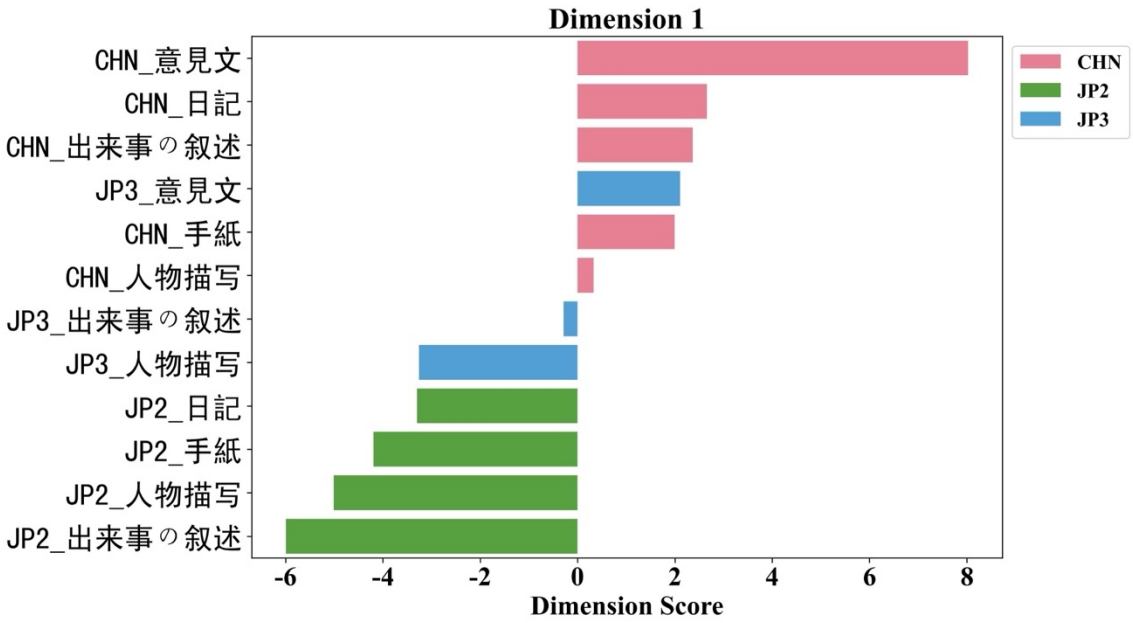

図 6.1　次元 1：産出語彙の豊富さ VS. 産出語彙の限定性

正の次元スコアが高いサブコーパス（例えば、CHN_意見文・CHN_日記）は、次元 1 の正次元に属する言語項目（例えば、非 HSK 語彙・高難度語彙）の出現頻度が高いのに対し、負の次元に属する「一人称単数代名詞“我”」や「低難度語彙」の頻度が著しく低い。逆に、負の次元スコア（絶対値）が高いジャンル（例えば、JP2_出来事の叙述・JP2_人物描写）は、正反対の言語項目を持っており、すなわち、「一人称単数代名詞“我”」や「低難度語彙」の頻度が非常に高く、「非 HSK 語彙」・「高難度語彙」などの頻度が明らかに少ないことを意味している。

本研究で特定した「低難度語彙」は HSK1～2 級の語彙、「中難度語彙」は HSK3～4 級の語彙、「高難度語彙」は HSK5～6 級の語彙、「非 HSK 語彙」は HSK1～6 級の語彙リストに収録されていない語彙である。これら 4 つの言語項目のうち、「低難度語彙」・「高難度語彙」・「非 HSK 語彙」という 3 つの言語項目は次元 1 に強い影響を与えていることが分かった。なお、一人称単数代名詞“我”は HSK 1 級レベルの語彙（低難度語彙）である。「韻律双音節語」は文章をより書面的に形成することができ、「語彙多様性」は産出した文章における語彙の豊富さを表す指標であり、「平均文長」は文の長さを測る指標である。

すなわち、次元 1 では、正のグループを構成する言語項目は主に産出語彙の豊富さを反映する言語項目であるのに対して、負のグループに属する言語項目は主に低難度語彙に限定していることであるといえる。従って、次元 1 は、「産出語彙の豊富さ VS. 産出語彙の限定性」というラベルが付けられると考えられる。つまり、次元 1 の Y 軸正方向に行けば行くほどジャンルの産出語彙の豊富さが高くなり、負方向に行けば行くほどジャンルの産出語彙の限定性が高くなる（もしくは、産出語彙の豊富さが低くなる）。

図 6.1 に示したジャンル間の類似点や相違点を見れば、次元 1 は学習者と母語話者が書いた作文間の区別をすることができる。なぜなら、中国語母語話者の作文ジャンルはいずれも正の次元に位置しているのに対し、日本人中国語学習者の作文ジャンルにおいて、「JP3_意見文」以外のどちらも負の次元に分布しているからである。図 6.1 は、次元 1 の機能という点では「CHN_意見文」の産出語彙の豊富さがもっとも顕著であるのに対し、「JP2_出来事の叙述」の産出語彙の限定性が一番低いことを示している。

下記の例文(4)～(6)それぞれは、「CHN_意見文」・「JP3_意見文」・「JP2_出来事の叙述」という 3 つのサブコーパスから抽出した作文の抜粋である。例文において、「低難度語彙」（例えば、是、的、*我*）は下線で示しており、「低難度語彙」である「一人

称単数代名詞“我”」は下線が引かれていることに加え、イタリック体で表示されている。「高難度語彙」（例、人类、爱护、充满）、「非 HSK 語彙」（例、感恩、呵护、安定）、「韻律双音節語」（例、结束、充满、安定）はそれぞれ、二重下線、波線、傍点で表示されている。

---

(4) 山清水秀， 鸟语花香， 风调雨顺， 五谷丰登， 是 人类 追求 向往 的 美 境，也 是 人类 劳动 创造 和 精心 爱护 的 硕果 。 水 在 不 停 地 运动， 在 人体 里， 在 农田， 在 工厂， 使 世界 充满 生机 和 活力， 当 我们 徜徉 在 大自然 的 怀抱 的 时候， 其实 我们 所 面对 的 全部 是 水 给 我们 的 力量 。 人类 体内， 百分之 七十 是 水 。 水 —— 给 了 我们 绿色 、 生命 和 希望，我们 有 什么 理由 不 对 它 心怀 感恩 并 珍惜 、 呵护 它 呢 ？

（CHN_意見文_081, Dimension 1 score: 8.18）

(5) *我* 喜欢 旅行， 但是 这样 的 情况 下， *我* 的 一 个 爱好 也 被 封口， 闷闷不乐 地 度过 每天 的 日子 。 *我* 最 难过 的 是 丢失 每天 的 学校 生活 的 机会， *我* 的 同班 同学 是 从 一 年级 一样 的， 上课 前 后 我们 本来 想 到 一起 吃饭， 一起 学习， 可是 上 网课 的 时候， 当然 我们 自己 的 旁边 儿 没有 这样 的 朋友 。 那， 什么 办法 可以 结束 这样 疫情 呢 ？ *我* 认为 最 好的 方法 是 接种 疫苗 。 这个 方法 不仅 有 医疗 性 的 效果， 也 有 精神性 的效果 。 *我* 有 两 个 理由 ： 群体 免疫 和 紧张 暖和 。

（JP3_意見文_023, Dimension 1 score: 1.19）

(6) 在 骑 自行车 的 时候， *我* 发现 路 上 躺 着 一 个 老人 。 *我* 以为 他 是 个醉鬼， 但 定神 一 瞧， 他 好像 是 昏厥 了 。 *我* 着急 地 走 上前 去 看 他，发现 他 果然 昏迷不醒 。 *我* 问 他 ： “ 你 好， 能 听 见 *我* 的 声音 吗 ？ ” 但 他 没有 反应 。 *我* 马上 叫 救护车， 但 不 知道 为什么， *我* 的 眼泪 止不 住 。 过 了 一会儿， 救护车 到达 了 。 一 名 医生 跟 *我* 说 ： “ 放心吧， 我们 一定 能 救活 他 。 ” 听 了 医生 的 话， *我* 心中 安定 了 一些 。 *我* 只有 为 他 祈祷 了 …

（JP2_出来事の叙述_002, Dimension 1 score: -6.20）

上記の例文から見られるように、母語話者の中国語作文は、全体的にセンテンスの長さが長く、語彙が豊富で、高難度語彙が多く使用しており、低難度語彙の出現頻度が低いことが分かった。 一方、学習者の中国語作文は、センテンスの長さが短く、一人称単数代名詞“我”のような低難度語彙を多用する傾向が見られた。

6.1.1 次元 2：状態状況表現の適切さ

表 6.3 は次元 2 の解釈に使用された言語項目しか表示していない。次元 2 では、正の言語項目は 4 つがあり、負の言語項目は 1 つがある。次元 2 を構成する言語項目において、「程度を表す副詞」が「ぼかし言葉」の一部であるため、負の言語項目は大きく分けて 3 種類に分類できる。つまり、ぼかし表現、接続詞、単語の平均長さである。正の言語項目である「時態助詞“着”」は、中国語アスペクト助詞の一種であり、主に動詞の後につき、「～している」のような動作や状態の持続を表す機能を持っている。

表 6.3　次元 2 に属する言語項目

Dimension 2	
時態助詞“着”	0.62
……	
単語の平均長さ	-0.52
接続詞	-0.60
程度を表す副詞	-0.75
ぼかし言葉	-0.84

図 6.2 は、次元 2 におけるジャンルの次元スコアの平均値を昇順に示した横棒グラフの可視化図である。図 6.2 に示したサブコーパス間の類似点や相違点を見れば、次元 2 の両極に分布しているのは学習者と母語話者の作文ジャンルであることを確認した。日本人中国語学習者の作文ジャンルがどちらも負の次元に分布しているのに対し、中国語母語話者の作文ジャンルはいずれも正の次元に位置している。そのため、次元 2 は中国語学習者と中国語母語話者が書いた作文を区別する次元であると言える。

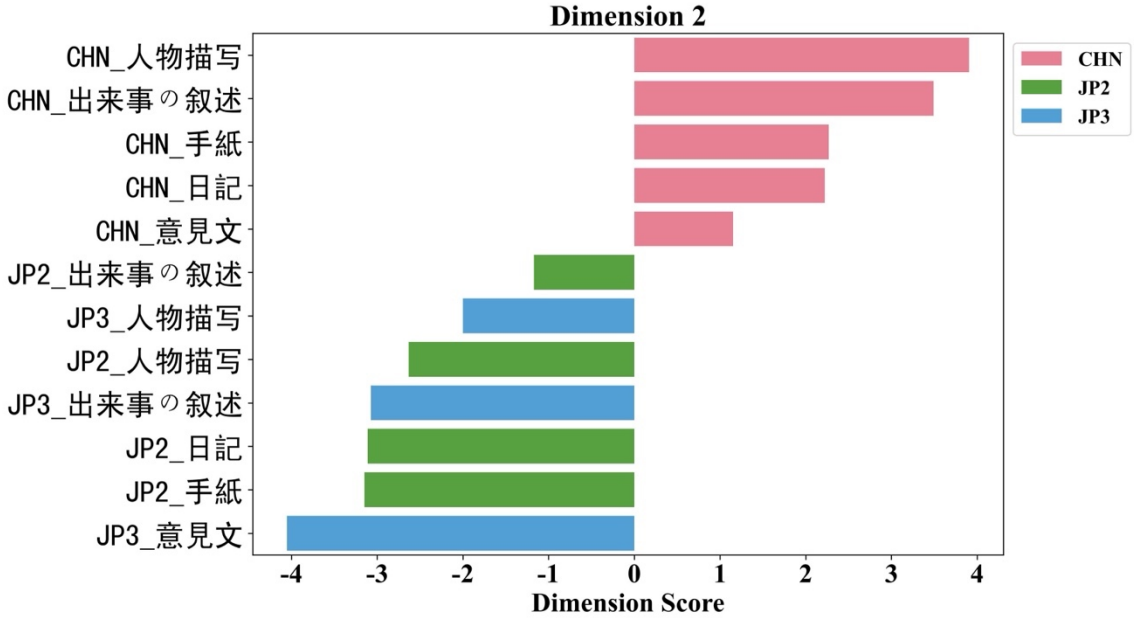

図 6.2　次元 2：状態状況表現の適切さ

下記の例文(7)～(9) はそれぞれ、「CHN_人物描写」・「JP3_意見文」・「JP3_出来事の叙述」という 3 つのジャンルから抽出した作文の抜粋である。例文において、「時態助詞“着”」は下線が引かれていることに加え、イタリック体で表示されている。「ぼかし言葉」（例えば、很、十分、很多）は波線で示しており、「接続詞」（例、但是、因为、如果）は二重下線で表示されている。

---

(7) 过年 的 时候 ， 家里 来 了 很多 人 ， 有 人 送 他 了 一 本 基本 笔画 练习 册 ， 他 十分 兴奋 ， 迫不及待 的 打开 练习 。 家里 人声 嘈杂 ， 大家 都 在 说 笑 ， 只有 弟弟 把 自己 “ 隔离 ” 了 起来 ， 眼睛 认真 的 盯 *着* ， 手 慢慢 的 写 *着* ， 认真 专注 ， 似乎 周围 很 安静 的 样子 。 我 特意 在 他 后面 放 了 一 个 计时器 ， 想 看看 他 能 坚持 多久 。 时间 一分一秒 的 过去 了 ， 弟弟 一 直 没有 停 下来 ， 直到 吃 午饭 ， 他 才 放 下 笔 ， 吃饭 后 又 去 练习 了 。

(CHN_人物描写_004, Dimension 2 score: 3.74)

(8) 众所周知 ， 新型 冠状 病毒 比 口罩 能够 遮挡 的 大小 还 小 。 但是 ， 人 打 喷嚏 、 咳嗽 时 扩散 的 飞沫 中 含有 很多 病毒 ， 口罩 能够 遮挡 这个 飞 沫 。

因此 ， 戴 口罩 对 防止 病毒 很 有效 。 然后 ， 我 介绍 一下 为什么 有的 人 拒绝 戴 口罩 。 我 听说 拒绝 戴 口罩 的 人 大多 是 年轻人 与 老年人 。 据 Yahoo 新闻 ， 走 在 *歌舞伎町* 的 人 的 百分之 十七 没 戴 口罩 。 接受 采访 的 *歌舞伎町* 的 陪酒哥 对 “ 为什么 你 没 带 口罩 ？ ” 这个 提问 这样 回答 ： “ 我 四月份 感染 了 病毒 ， 已经 有 抗体 了 ， 所以 不 用 带 口罩 。 ” 别 的 陪酒 哥 又 这样 回答 ： “ 新型 冠状病毒 是 未知 的 ， 戴 口罩 也 没有 效果 ， 太 麻烦 了 。 ” 老年人 的 回答 与 他们 的 回答 很 相似 ： “ 因为 我 很 健康 ， 病毒 跟 我 完全 无关 。 “ 这些 回答 都 是 很 自私 的 ， 没有 考虑 别人 。 我 希望 所有 人 为了 自己 、 别人 与 社会 戴 好 口罩 。

（JP3_意見文_005, Dimension 2 score: -4.06）

(9) 通过 这个 上海 留学 ， 我 觉得 中国 人 很 像 大阪人 。 即使 不 认识 人 也 会 和 我们 搭话 。 “ 樱花 在 哪 可以 看 ？ ” “ 你 的 中文 不错 。 ” 让 我 吃惊 场面 很多 ， 但 我 很 高兴 因为 知道 我 可以 跟 中国 人 聊天 。 我 很 喜欢 中国 人 外向 性格 。 如果 没有 新冠 病毒 ， 今年九月 我 会 去 中国 留学 。 但是 现在 我 打算 从 明年二月 去 浙江 留学 。 非常 希望 能够 实现 这个 留学 梦想 。

（JP3_出来事の叙述_041, Dimension 2 score: -3.51）

上記の例文が示したように、母語話者に比べて学習者は、作文における難度の低いぼかし表現（例、“很”/“非常”/“太”）と接続詞（例、“所以”/“但是”/“因为”）の出現頻度が高く、単語の長さの平均値が（例えば“*歌舞伎町*”、日本の人名・地名が長いことを反映して）全体的に長い傾向が見られ、一方で助詞“着”（Verb +“着”：～している、動作や状態の持続を表す）のような状態状況表現はあまり使用されていないことが分かった。そのため、次元 2 には「状態状況表現の適切さ」という解釈ラベルを付けることができるであろう。学習者の作文における状態状況表現の適切さが低い原因は、作文を書く時、学習者は、主に勉強や生活で使い慣れた語彙を多く使い、母語話者のようにジャンルに応じた適切な状態状況表現を使うまでには至っていないと推察される。

6.1.3 次元 3：副詞の修飾

次元 3 に属する「単語の平均長さ」は次元 1 と次元 2 に重複して現れており、次元 2 における負荷量の方が高いため、この項目は次元 3 の解釈に使用されない。すなわち、次元 3 をラベリングする際、主に正の負荷量を持っている言語項目「高頻度副詞」・「時間を表す副詞」で解釈を行う。表 6.4 は次元 3 の解釈に使用する言語項目しか表示していない。次元 3 では副詞、特に時間を表す副詞の使用が特徴的であったため、次元 3 には「副詞の修飾」という解釈ラベルが付けられる。

表 6.4　次元3に属する言語項目

Dimension 3	
高頻度副詞	1.17
時間を表す副詞	0.71

図 6.3 に示したジャンル間の類似点や相違点を見れば、中国語母語話者の作文ジャンルにおいて、「CHN_意見文」以外のいずれも正の次元に位置している。それに対し、日本人中国語学習者の作文ジャンルにおいて、「JP3_人物描写」以外のどちらも負の次元に分布している。そのため、次元 3 もほぼ学習者と母語話者の作文を区別する次元であると言える。

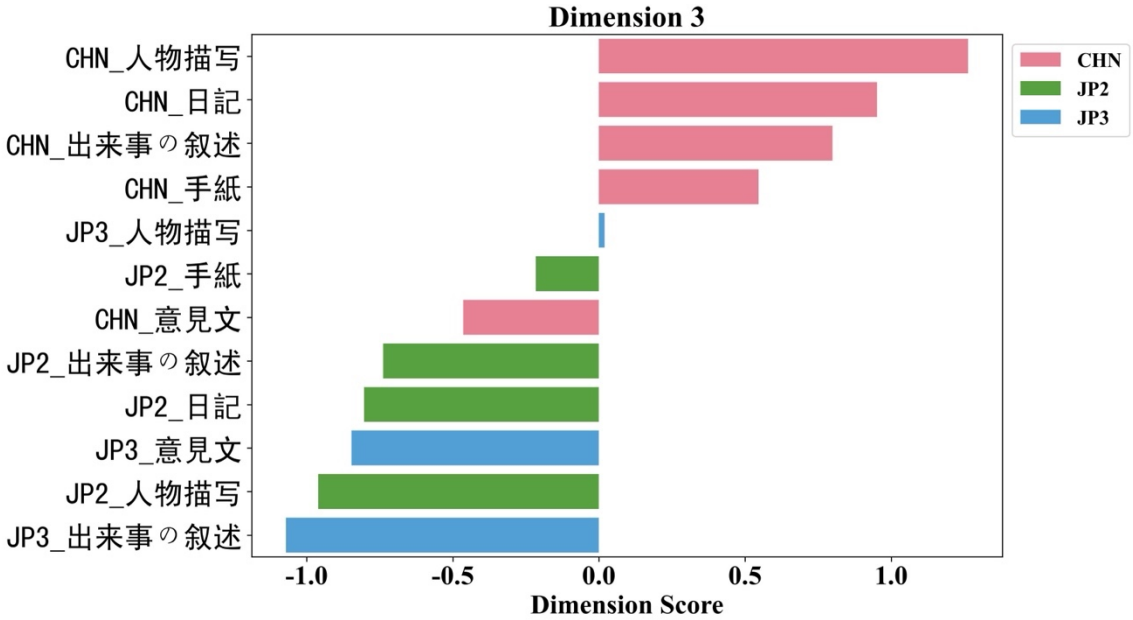

図 6.3　次元 3：副詞の修飾

下記の例文(10)～(13) は、「出来事の叙述」・「人物描写」のジャンルから抽出した作文の抜粋である。例文において、「高頻度副詞」に属する語彙に下線が引かれており、「時間を表す副詞」は傍点を付けられている。

---

(10) 要 说 我 爸爸 ， 他 真 的 是 一个 名副其实 的 球迷 ， 特别 是 对 篮球赛 情有独钟 。 而 他 最 爱 看 的 球赛 又 总是 在 晚上 11 点 和 12 点 播出 ， 所以 他 经常 卷起 铺盖 跑 到 客厅 看 球 ， 以 解 自己 的 “ 球瘾 ” ， 又 不 影响 我们 的 正常 休息 。 每 一 次 ， 他 总是 早早 地 躺 在 沙发 上 ， 准时 准点 地 打开 电视 ， 等到 球赛 开始 ， 他 便 津津有味 地 看 起 篮球赛 了 。 每当 他 看 好 的 球队 失手 的 时候 ， 他 总是 眼睛 睁 得 好 大 ， 眉头 紧锁 ， 恨不得 跑 进 篮球场 上 去 给 他们 当 总 指挥 ， 每当 球队 进球 的 时候 ， 他 就 好像 条件反射 一样 地 跳 了 起来 ， 嘴角 向上 微微 翘 起 ， 眼睛 微 眯 ， 露出 了 月牙 般 的 弧度 ， 然后 大声 地 说 ： “ 哈哈哈 ， 进球 啦 ， 真 不愧 是 我 看 好 的 球队 啊 ！ ” 当 他 看 好 的 球队 与 别的队 只 剩 一 球 的 分 差 时 ， 他 便 会 将 眼睛 和 嘴巴 忽然 张开 ， 眉毛 上扬 ， 就 连 鼻孔 都 比 平时 大 了 许多 。 高兴 地 说 ： “ 哈哈哈 ， 我们 赢 啦 ， 我 看 好 的 球队 太 厉害 啦 ！ ”

（CHN_人物描写_005, Dimension 3 score: 1.64）

(11) 真 没 想到 包 饺子 会 这么 难 。 我 不禁 有些 烦 了 ， 不 想 包 了 。 可 转念 一 想 ， 我 刚才 还 满 口 答应 过 妈妈 呢 ， 难道 要 临阵脱逃 ？ 看来 只能 拜师 学艺 了 。 我 认真 地 观察 起来 ， 只见 妈妈 拿 了 一个 皮 ， 告诉 我 说 ： “ 放 馅 的 时候 要 适量 ， 过 多 会 溢 出来 ， 过 少 会 不 好吃 。 ” 接着 她 用 双手 的 拇指 和 食指 分别 捏 着 饺子 的 左右 两边 ， 往中间 轻轻 一 捏 ， 一个 挺 着 将军肚 的 饺子 就 成型 了 ， 看上去 还 挺 英俊 的 。 接下来 ， 我 在 妈妈 的 嘱咐 和 帮助 下 ， 我 恢复 了 自信 ， 再 一 次 开始 包 了 。 这 回 我 抓住 了 包 饺子 的 重点 ， 馅 放 得 不 多 也 不 少 ， 对折 时 力量 也 适中 ， 左右手 配合 默契 ， 同时 往 中间 捏 。 我

小心翼翼 地 操作 着 每 个 步骤 。 功夫 不 负 有心 人 ， 我 的 第一 个 饺子 终于 “ 出世 ” 了 。 （CHN_出来事の叙述_038, Dimension 3 score: 0.39）

(12) 她 虽然 学习 成绩 不 好 ， 但是 我 觉得 她 是 想 出 新 玩儿 的 天才 ， 每天 她 产生 的 各种各样 的 玩儿 让 人 很 开心 。 我 觉得 这 是 一个 她 的 才能 之一 。 再说 ， 我们 乒乓球 社团 的 练习 太 辛苦 了 ， 但是 我 能 没 放弃 地 坚持 ， 那 真是 多亏 了 她 的 存在 ， 如果 没有 她 的 幽默感 ， 我们 会 从 辛苦 的 练习 逃避 的 。 她 现在 住 在 东京 ， 是 个 美容 专科学校 的 学生 ， 从 明年 春天 开始 在 化妆品 店 工作 。 我 想 去 东京 访问 她 的 家 。

（JP3_人物描写_059, Dimension 3 score: 0.46）

(13) 演剧 演 完了 ， 我 还有 时间 到 夜班车 出发 为止 。 所以 我 去 了 涩谷 。 涩谷 叫做 “ 年轻人 的 城市 ” ， 有 很多 购物 中心 。 我 在 那儿 高兴 地 浏览 商店 橱窗 ， 在 咖啡厅 喝 了 一 杯 咖啡 和 休息 会儿 。 全 一 天 我 活动 了 ， 所以 从 东京 坐 夜班车 回 大阪 的 时候 ， 我 会 睡 得 很 熟 。这个 东京 之 行 非常 非常 开心 。 （JP3_出来事の叙述_052, Dimension 3 score: -0.84）

例文の抜粋と図 6.3 を合わせて見ると、母語話者作文において、副詞、特に時間を表す副詞（例、经常、总是、忽然）が頻繁に使用されているのに対し、学習者作文ジャンルの多くでは、副詞、特に「時間を表す副詞」があまり用いられていないことが確認された。

6.1.4 次元 4：動作の描写

次元 4 には、3 つの正の言語項目のみが含まれている。次元 4 をラベリングする際、主に正の負荷量を持っている言語項目「動作動詞」・「方向動詞」・「高頻度動詞」で解釈を行う。表 6.5 から分かるように、次元 4 では動詞の使用が特徴的であったため、次元 4 には「動作の描写」という機能ラベルが付与される。

表6.5　次元4に属する言語項目

Dimension 4	
動作動詞	0.92
方向動詞	0.86
高頻度動詞	0.54

また、図 6.4 に示すように、「出来事の叙述」・「人物描写」のようなジャンルでは動詞の出現頻度が高いが、「意見文」における動詞の出現頻度が低い。従って、この次元は作文ジャンル間の区別を反映していると言えるであろう。

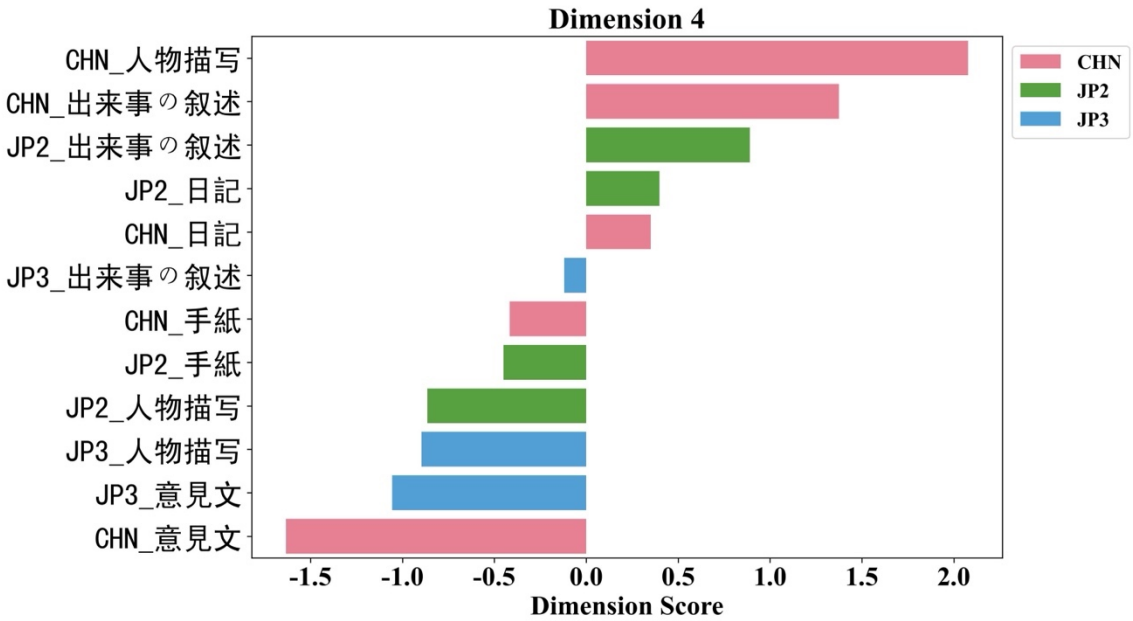

図 6.4　次元 4：動作の描写

下記の例文(14)～(17) は、「意見文」と「人物描写」のジャンルから抽出した作文の抜粋である。例文において、「高頻度動詞」には下線が引かれ（例：“<u>发现</u>”、“<u>参加</u>”）、「動作動詞」に傍点を付け（例：“走”、“捧”）、「方向動詞」はイタリック体（例：“*过来*”、“*起来*”）で表示されている。

---

(14) 奶奶 发现 我 站 在 门口 ， 她 赶紧 放下 手中 的 碗筷 走 *过来* ， 看到 我 捧 着 的 空 碗 ， 她 好像 松 一 口 气 ， 布满 皱纹 的 脸上 露出 了 开心 的 笑 容 。 她 冲 我 拍拍 自己 的 肚子 ， 又 指 了 指 我 手中 的 碗 ， 好像 在 用 老家 话 问 我 吃 饱 了 没有 。 那 下 子 ， 我 不禁 一下子 听 懂 了 她 的 意 思 ， 听 出 了 她 话 中 浓浓的 关切 之 情 。

（CHN_人物描写_051, Dimension 4 score: 2.16）

(15) 她 是 一个 中国 女 孩儿 ， 日语 和 汉语 都 会 说 ， 所谓 的 双语者 。 因为 她 从 小 到 现在 在 日本 长大 了 ， 一直 在 日本 学校 学习 ， 还有 她 的 爸爸 和 妈妈 都 是 中国 人 ， 在 家里 经常 说 汉语 。 她 长 得 很 漂亮 ， 眼睛 又 大 又 明亮 ， 睫毛 非常 长 ， 牙齿 很 整齐 ， 笑 *起来* 很 好看 。

（JP3_人物描写_013, Dimension 4 score: -0.62）

(16) 在 日本 ， 有些 人 也 参加 不 戴 口罩 的 运动 ， 但是 大多数 人 反对 那儿 的 想法 。 有些 人 觉得 口罩 侵犯 了 他们 的 选择 自由 ， 但 如果 更 多 的 人 戴 上 ， 我们 将 有 更 多 的 自由 外出 。 不光 为了 自己 ， 也 为了 自己 的 家人 和 朋友 ， 我们 要 戴 口罩 。

（JP3_意見文_006, Dimension 4 score: -1.42）

(17) 短短 几 天 时间 ， 意大利 首席 医务官 因 新冠 病毒 去世 ， 加拿大 总理 妻子 确诊 ， 伊朗 死亡率 居高不下 ， NBA 球员 、 奥斯卡 影帝 接连 被 确诊 …… 病毒 的 蔓延 超过 了 所有 人 的 预料 ， 正如 世卫组织 3 月 11 日 所 言 ， 新冠 病毒 疫情 已 发展 为 大 流行病 。

（CHN_意見文_116, Dimension 4 score: -2.58）

例文の抜粋と図 6.4 を合わせて見ると、「人物描写」のような作文ジャンルは、動詞、特に動作動詞（例、“站”/“走”/“吃”）と方向動詞（例、“过来”/“起来”）がよく使用されているのに対し、「意見文」のような作文ジャンルは、動作動詞や方向動詞があまり用いられていないことが分かった。

6.2 同一次元におけるジャンル間の比較

本節では、次元スコアの Radar chart（レーダーチャート） と One-way ANOVA の多重比較を用い、母語話者作文におけるジャンル間の比較と同一作文ジャンルにおける母語話者作文と学習者作文の比較を行う。

6.2.1 母語話者作文におけるジャンル間の比較

まず、母語話者によって書かれた作文ジャンル間の比較を行う。表 6.6 に母語話者作文ジャンルに関する各次元スコアの平均値を示している。図 6.5 は表 6.6 に基づいて作成しており、母語話者作文におけるジャンル間の比較を可視化した Radar chart[73]図である。

表6.6　母語話者作文ジャンルに関する各次元スコアの平均値

Corpus	Subcorpus	N	次元 1	次元 2	次元 3	次元 4
中国語母語話者	CHN_日記	136	2.66	2.22	0.95	0.35
	CHN_出来事の叙述	133	2.37	3.49	0.80	1.38
	CHN_人物描写	137	0.33	3.91	1.26	2.08
	CHN_手紙	136	2.00	2.27	0.55	-0.42
	CHN_意見文	132	8.03	1.15	-0.46	-1.63

[73] Radar chart（日本語訳：レーダーチャート）は、スパイダーチャートとも呼ばれ、複数の項目（変数）を構成比に変換せずにまとめて比較する方法である。

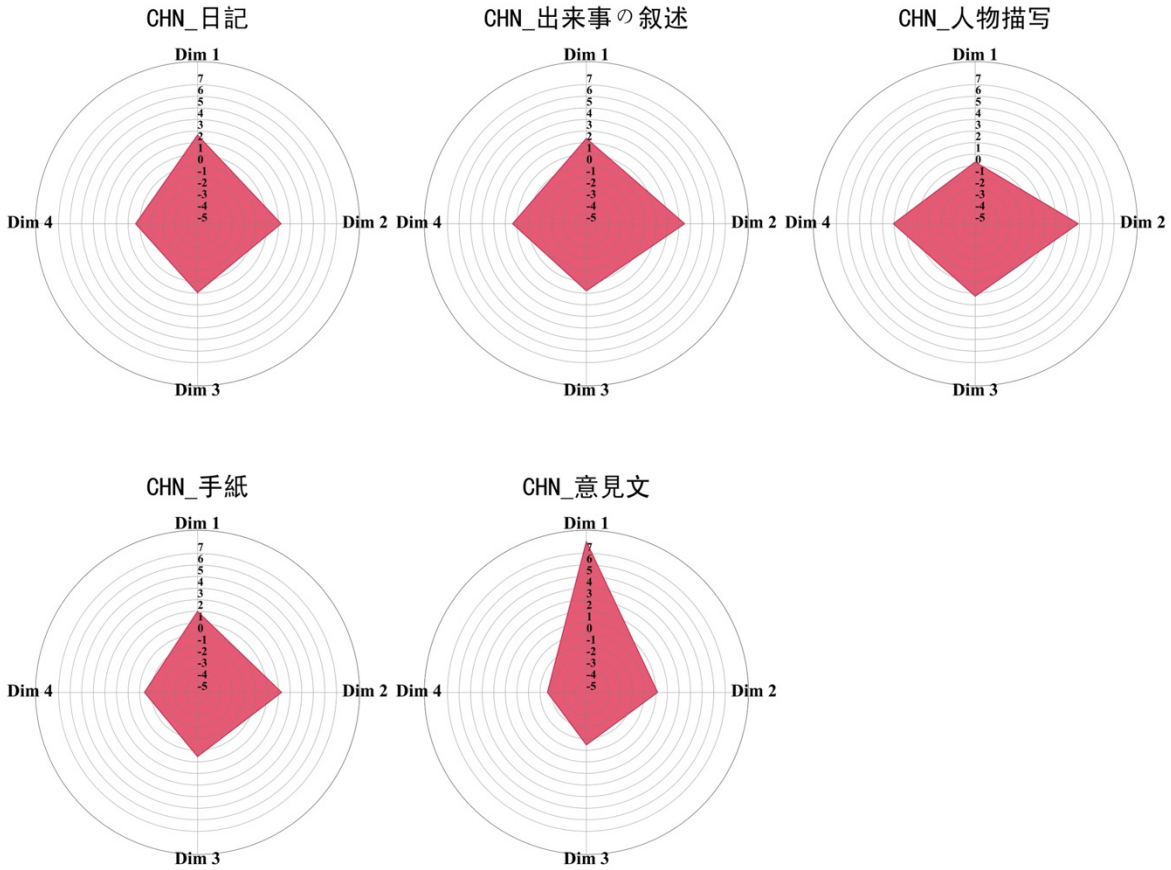

図 6.5　母語話者作文ジャンル間の比較（radar chart）

表 6.6 の平均値表と図 6.5 の可視化図を組み合わせて見ると、各次元で各ジャンルにばらつきがあることがわかった。 全体として、ほかの 4 つのジャンル（日記・出来事の叙述・人物描写・手紙）に比べて、「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。また、5 つの作文ジャンルは各次元での平均値の差が統計的に有意であるかどうかを確認するために、One-way ANOVA 分析を行った。表 6.7 は One-way ANOVA 分析[74]を行った多重比較（Tukey's HSD）の結果である。

多重比較の結果によれば、次元 1 「産出語彙の豊富さ VS. 産出語彙の限定性」について、「意見文」と他の 4 つのジャンル間に有意差があり（$p < 0.05$）、「日記」・「出来事の叙述」・「手紙」という 3 つのジャンル間には有意差がなく（$p > 0.05$）、「人物描写」と他の 4 つのジャンル間に有意差がある（$p < 0.05$）ことが明らかになった。ほかのジャンルに比べて、「意見文」の産出語彙が最も豊富であるのに対し、「人物描写」の

[74] Python での One-way ANOVA 分析は次のリンクを参考にした。
https://www.reneshbedre.com/blog/anova.html〔最終アクセス日：2022 年 9 月 25 日〕

産出語彙の豊富さが一番低いことが確認された。産出語彙の豊富さは高い順に、「意見文＞日記≈ 出来事の叙述 ≈ 手紙 ＞人物描写」となっている。

表6.7　母語話者作文ジャンルに関する各次元スコアの多重比較（Tukey's HSD）

		次元 1		次元 2		次元 3		次元 4	
Group1	**Group2**	**Diff**	**Sig.**	**Diff**	**Sig.**	**Diff**	**Sig.**	**Diff**	**Sig.**
手紙	人物描写	1.666	0.001	-1.641	0.001	-0.717	0.004	-2.495	0.001
手紙	出来事の叙述	-0.371	0.900	-1.225	0.001	-0.252	0.706	-1.792	0.001
手紙	日記	-0.663	0.549	0.043	0.900	-0.405	0.273	-0.768	0.075
手紙	意見文	-6.031	0.001	1.113	0.001	1.011	0.001	1.218	0.001
人物描写	出来事の叙述	-2.037	0.001	0.416	0.510	0.464	0.156	0.702	0.130
人物描写	日記	-2.330	0.001	1.684	0.001	0.312	0.535	1.726	0.001
人物描写	意見文	-7.697	0.001	2.754	0.001	1.728	0.001	3.713	0.001
出来事の叙述	日記	-0.293	0.900	1.268	0.001	-0.152	0.900	1.024	0.006
出来事の叙述	意見文	-5.660	0.001	2.338	0.001	1.264	0.001	3.010	0.001
日記	意見文	-5.368	0.001	1.070	0.001	1.416	0.001	1.986	0.001

注：「Diff」は、Group1 と Group2 の平均値の差であり、「Sig.」は *p*-value である。

次元 2「状態状況表現の適切さ」について、「人物描写」と「出来事の叙述」の間に有意差がなく（$p > 0.05$）、「人物描写 / 出来事の叙述」と他の 3 つのジャンル間に有意差があり（$p < 0.05$）、「手紙」と「日記」の間に有意差がなく（$p > 0.05$）、「手紙 / 日記」と「意見文」の間に有意差がある（$p < 0.05$）ことを確認した。ほかの 3 つジャンルに比べて、「人物描写」と「出来事の叙述」における状態状況表現が最も多いのに対し、「意見文」おける状態状況表現が一番少ないことが分かった。状態状況表現の適切さは高い順に、「出来事の叙述 ≈ 人物描写 > 日記 ≈ 手紙 > 意見文」となっている。

次元 3「副詞の修飾」について、「人物描写」・「日記」・「出来事の叙述」の間に有意差がなく（$p > 0.05$）、「意見文」と他の 4 つのジャンル間に有意差がある（$p < 0.05$）ことが確認された。しかし、「手紙」は「人物描写」や「意見文」との間に有意差がある（$p < 0.05$）が、「日記」や「出来事の叙述」との間に有意差がない（$p > 0.05$）。その原因で、副詞の修飾について、5 つのジャンルを高い順に並べるのが難しいため、「手紙」というジャンルを除いて他の 4 つのジャンルで並べてみると、副詞の

修飾は高い順に、「人物描写 ≈ 日記 ≈ 出来事の叙述 ＞意見文」となっている。つまり、「意見文」における副詞の修飾は一番少ないことが分かった。

次元 4「動作の描写」について、「人物描写」と「出来事の叙述」の間に有意差がない（$p > 0.05$）のに対し、「人物描写 / 出来事の叙述」と他に 3 つのジャンル間に有意差がある（$p < 0.05$）ことを明らかにした。また、「日記」と「手紙」の間に有意差がなく（$p > 0.05$）、「意見文」と他の 4 つのジャンル間に有意差がある（$p < 0.05$）ことも確認した。この結果から分かるように、「人物描写」や「出来事の叙述」における動作の描写が最も高いのに対し、意見文における動作の描写が一番少ない。つまり、動作の描写は高い順に、「人物描写 ≈ 出来事の叙述 ＞日記 ≈ 手紙＞意見文」となっている。

表 6.8 母語話者の作文ジャンル：多重比較の結果のまとめ

次元	多重比較の結果
次元 1： 産出語彙の豊富さ vs. 産出語彙の限定性	意見文 ＞日記≈出来事の叙述≈手紙＞人物描写
次元 2：状態状況表現の適切さ	出来事の叙述≈人物描写＞日記 ≈手紙＞意見文
次元 3： 副詞の修飾	人物描写≈日記≈出来事の叙述 ＞意見文
次元 4： 動作の描写	人物描写≈出来事の叙述 ＞日記≈手紙＞意見文

注：「＞」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

表 6.8 は、母語話者の作文ジャンルにおける多重比較の結果をまとめたものである。全体から明らかなように、「意見文」以外の他の 4 つのジャンルは、全体的にある程度の共通性を持っている傾向がある。例えば、「出来事の叙述」と 「人物描写」では、産出語彙の豊富さ、状態状況表現、動作の描写などの特徴が比較的似ている。

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。

6.2.2 同一作文ジャンルにおける母語話者作文と学習者作文の比較

（1）ジャンル：人物描写

表 6.9 に「人物描写」という作文ジャンルに関する各次元スコアの平均値を示している。図 6.6 は表 6.9 に基づいて作成ており、作文ジャンル「人物描写」における母語話者作文と学習者作文の比較を可視化した Radar chart 図である。

表6.9　「人物描写」に関する各次元スコアの平均値

Subcorpus	N	次元 1	次元 2	次元 3	次元 4
CHN_人物描写	137	0.33	3.91	1.26	2.08
JP2_人物描写	163	-5.01	-2.63	-0.96	-0.86
JP3_人物描写	103	-3.26	-2.00	0.02	-0.90

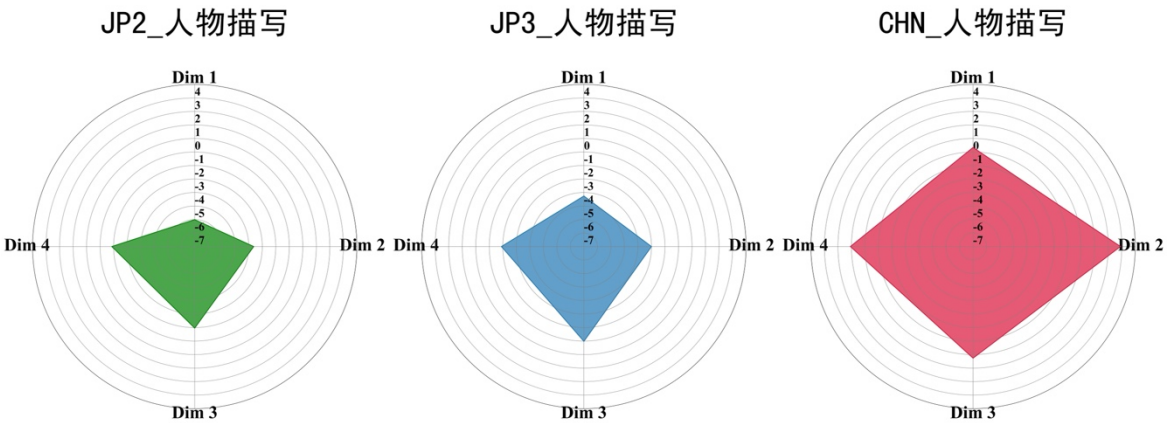

図 6.6　「人物描写」：母語話者作文と学習者作文の比較（radar chart）

表 6.9 の平均値表と図 6.6 の可視化図から明らかなように、母語話者と学習者によって書かれた「人物描写」の作文には、ばらつきがあり、母語話者の方が学習者より全体的に各次元のスコアが高い傾向が見られる。母語話者と学習者の「人物描写」作文ジャンルは各次元での平均値の差が統計的に有意であるかどうかを確かめるために、One-way ANOVA 分析を行った。表 6.10 は One-way ANOVA 分析を行った多重比較の結果である。

表6.10　「人物描写」に関する各次元スコアの多重比較（Tukey's HSD）

		次元 1		次元 2		次元 3		次元 4	
Group1	**Group2**	**Diff**	**Sig.**	**Diff**	**Sig.**	**Diff**	**Sig.**	**Diff**	**Sig.**
CHN_人物描写	JP2_人物描写	5.343	0.001	6.541	0.001	2.224	0.001	2.943	0.001
CHN_人物描写	JP3_人物描写	3.592	0.001	5.907	0.001	1.244	0.001	2.974	0.001
JP2_人物描写	JP3_人物描写	-1.751	0.001	-0.634	0.074	-0.980	0.001	0.031	0.900

注：「Diff」は、Group1 と Group2 の平均値の差であり、「Sig.」は *p*-value である。

表 6.10 によると、次元 1 「産出語彙の豊富さ VS. 産出語彙の限定性」と次元 3「副詞の修飾」は CHN > JP3 > JP2 の順で 、学習者（JP2、JP3）と母語話者（CHN）の間（次元 1：$p = 0.001 < 0.05$, $p = 0.001 < 0.05$; 次元 3：$p = 0.001 < 0.05$, $p = 0.001 < 0.05$）、及び学習者（JP2 と JP3）の間（次元 1：$p = 0.001 < 0.05$; 次元 3：$p = 0.001 < 0.05$）において両方とも有意差があることが明らかになった。学習者は上級（JP3）になると、「人物描写」の作文における産出語彙の豊富さと副詞の修飾が増える傾向が見られるものの、学習者の方が母語話者より全体的に産出語彙の豊富さが低く、副詞の修飾が少ない。

次元 2「状態状況表現の適切さ」と次元 4「動作の描写」について、学習者（JP2 と JP3）の間に有意差が認めないものの（次元 2：$p = 0.074 > 0.05$; 次元 4：$p = 0.900 > 0.05$）、学習者（JP2、JP3）と母語話者（CHN）の間（次元 2：$p = 0.001 < 0.05$, $p = 0.001 < 0.05$; 次元 4：$p = 0.001 < 0.05$, $p = 0.001 < 0.05$）に有意差があり、人物描写の作文を書く際、学習者の方が状態状況表現や動作の描写を過少に使用することが分かった。

(2) ジャンル：出来事の叙述

表 6.11 に「出来事の叙述」のジャンルに関する各次元スコアの平均値を示している。図 6.7 は母語話者作文と学習者作文の比較を可視化した Radar chart 図である。

表6.11　「出来事の叙述」に関する各次元スコアの平均値

Subcorpus	**N**	次元 1	次元 2	次元 3	次元 4
CHN_出来事の叙述	133	2.37	3.49	0.80	1.38
JP2_出来事の叙述	40	-6.00	-1.17	-0.74	0.89
JP3_出来事の叙述	59	-0.29	-3.07	-1.07	-0.12

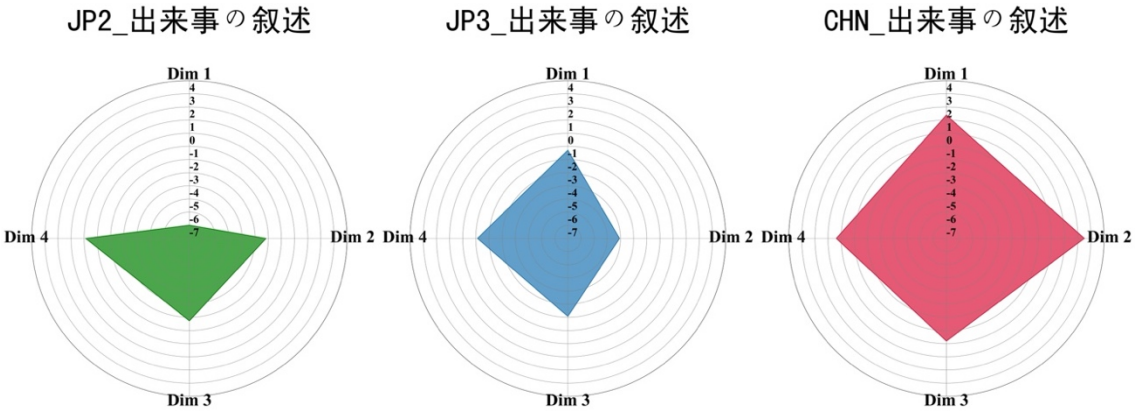

図 6.7　「出来事の叙述」：母語話者作文と学習者作文の比較（radar chart）

表 6.11 の平均値表と図 6.7 の可視化図から分かるように、母語話者と学習者の「出来事の叙述」の作文には、ばらつきがあり、母語話者の方が学習者より全体的に各次元のスコアが高い傾向が見られる。母語話者と学習者の「出来事の叙述」作文ジャンルは各次元での平均値の差が統計的に有意であるかどうかを確かめるために、One-way ANOVA 分析を行った。表 6.12 は One-way ANOVA 分析を行った多重比較の結果である。

表6.12　「出来事の叙述」に関する各次元スコアの多重比較（Tukey's HSD）

		次元 1		次元 2		次元 3		次元 4	
Group1	**Group2**	**Diff**	**Sig.**	**Diff**	**Sig.**	**Diff**	**Sig.**	**Diff**	**Sig.**
CHN_出来事の叙述	JP2_出来事の叙述	8.365	0.001	4.663	0.001	1.537	0.001	0.486	0.542
CHN_出来事の叙述	JP3_出来事の叙述	2.657	0.001	6.564	0.001	1.870	0.001	1.496	0.001
JP2_出来事の叙述	JP3_出来事の叙述	-5.708	0.001	1.901	0.001	0.333	0.576	1.010	0.133

注：「Diff」は、Group1 と Group2 の平均値の差であり、「Sig.」は *p*-value である。

表 6.12 に示した結果によれば、次元 1 「産出語彙の豊富さ VS. 産出語彙の限定性」は CHN > JP3 > JP2 の順、次元 2「状態状況表現の適切さ」は CHN > JP2 > JP3 の順で 、学習者（JP2、JP3）と母語話者（CHN）の間（次元 1：$p = 0.001 < 0.05$, $p = 0.001 < 0.05$; 次元 2：$p = 0.001 < 0.05$, $p = 0.001 < 0.05$）、及び学習者（JP2 と JP3）の間（次元 1：$p = 0.001 < 0.05$; 次元 2：$p = 0.001 < 0.05$）において両方とも有意差があることが分かった。

次元 3「副詞の修飾」は、学習者（JP2 と JP3）の間に有意差が認めないものの（p = 0.576 > 0.05）、学習者（JP2、JP3）と母語話者（CHN）の間（p = 0.001 < 0.05, p = 0.001 < 0.05）に有意差があり、人物描写の作文を書く際、学習者の方が副詞の修飾を過少に使用することが明らかになった。

次元 4「動作の描写」について、学習者（JP2）と母語話者（CHN）の間（p = 0.542 > 0.05）、および学習者（JP2 と JP3）の間（p = 0.133 > 0.05）に有意差は認められないものの、学習者（JP3）と母語話者（CHN）の間（p = 0.001 < 0.05）に有意差があることが分かった。

(3) ジャンル：手紙

表 6.13 は、「手紙」という作文ジャンルに関する各次元スコアの平均値を示している。図 6.8 は、「手紙」というジャンルにおける母語話者作文と学習者作文の比較を可視化した Radar chart 図である。

表6.13　「手紙」に関する各次元スコアの平均値

Subcorpus	N	次元 1	次元 2	次元 3	次元 4
CHN_手紙	136	2.00	2.27	0.55	-0.42
JP2_手紙	74	-4.20	-3.15	-0.22	-0.45

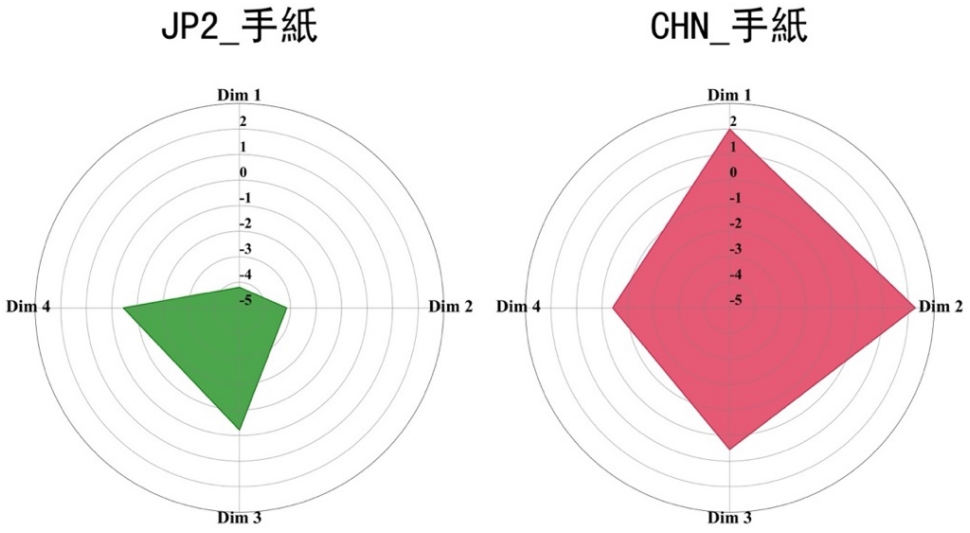

図 6.8　「手紙」：母語話者作文と学習者作文の比較（radar chart）

表 6.13 の平均値表と図 6.8 の可視化図から分かるように、母語話者と学習者によって書かれた「手紙」の作文には、ばらつきがあり、母語話者の方が学習者より全体的に各次元のスコアが高い傾向が見られる。

表6.14　「手紙」に関する各次元スコアの多重比較

		次元 1		次元 2		次元 3		次元 4	
Group1	**Group2**	**Diff**	**Sig.**	**Diff**	**Sig.**	**Diff**	**Sig.**	**Diff**	**Sig.**
CHN_手紙	JP2_手紙	6.196	0.001	5.416	0.001	0.762	0.001	0.034	0.900

注：「Diff」は、Group1 と Group2 の平均値の差であり、「Sig.」は *p*-value である。

母語話者と学習者によって書かれた「手紙」作文は各次元での平均値の差が統計的に有意であるかどうかを確かめた結果（表 6.14）、次元 4「動作の描写」について、学習者（JP2）と母語話者（CHN）の間に有意差は認められないものの（$p = 0.900 > 0.05$）、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」は、CHN > JP2 の順で 、学習者（JP2）と母語話者（CHN）の間（次元 1：$p = 0.001 < 0.05$; 次元 2：$p = 0.001 < 0.05$; 次元 3：$p = 0.001 < 0.05$）に有意差があることを確認した。手紙の作文を書く際、学習者（JP2）の方が産出語彙の豊富さと状態状況表現の適切さが低く、副詞の修飾を過少に使用することが分かった。

(4) ジャンル：日記

表 6.15 は、「日記」という作文ジャンルに関する各次元スコアの平均値を示している。図 6.9 は表 6.15 に基づいて作成ており、「日記」という作文ジャンルにおける母語話者作文と学習者作文の比較を可視化した Radar chart 図である。

表6.15　「日記」に関する各次元スコアの平均値

Subcorpus	**N**	次元 1	次元 2	次元 3	次元 4
CHN_日記	136	2.66	2.22	0.95	0.35
JP2_日記	142	-3.30	-3.11	-0.80	0.40

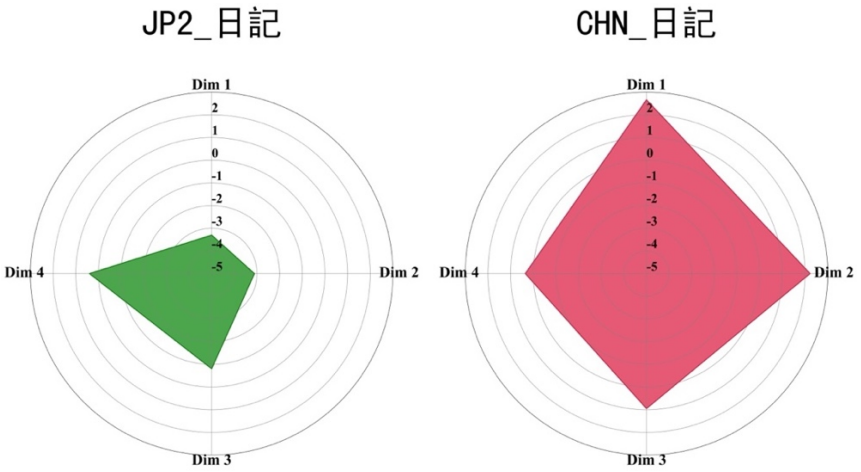

図 6.9　「日記」：母語話者作文と学習者作文の比較（radar chart）

表 6.15 の平均値表と図 6.9 の可視化図から分かるように、母語話者と学習者によって書かれた「日記」の作文には、ばらつきがあり、母語話者の方が学習者より全体的に各次元のスコアが高い傾向が見られる。

表6.16　「日記」に関する各次元スコアの多重比較

		次元 1		次元 2		次元 3		次元 4	
Group1	**Group2**	**Diff**	**Sig.**	**Diff**	**Sig.**	**Diff**	**Sig.**	**Diff**	**Sig.**
CHN_日記	JP2_日記	5.963	0.001	5.334	0.001	1.755	0.001	-0.047	0.873

注：「Diff」は、Group1 と Group2 の平均値の差であり、「Sig.」は *p*-value である。

母語話者と学習者の「日記」作文ジャンルは各次元での平均値の差が統計的に有意であるかどうかを確かめた結果（表 6.16）、次元 4「動作の描写」について、学習者（JP2）と母語話者（CHN）の間に有意差は認められないものの（$p = 0.873 > 0.05$）、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」は、CHN > JP2 の順で 、学習者（JP2）と母語話者（CHN）の間（次元 1：$p = 0.001 < 0.05$; 次元 2：$p = 0.001 < 0.05$; 次元 3：$p = 0.001 < 0.05$）に有意差があることを確認した。日記のような作文を書く際、学習者（JP2）の方が産出語彙の豊富さと状態状況表現の適切さが低く、副詞の修飾を過少に使用することが明らかになった。

(5) ジャンル：意見文

表 6.17 は、「意見文」という作文ジャンルに関する各次元スコアの平均値を示している。図 6.10 は表 6.17 に基づいて作成した、作文ジャンル「意見文」における母語話者作文と学習者作文の比較を可視化した Radar chart 図である。表 6.17 と図 6.10 から分かるように、母語話者と学習者によって書かれた「意見文」の作文には、ばらつきがあり、母語話者の方が学習者より全体的に各次元のスコアが高い傾向が見られる。

表6.17　「意見文」に関する各次元スコアの平均値

Subcorpus	N	次元 1	次元 2	次元 3	次元 4
CHN_意見文	132	8.03	1.15	-0.46	-1.63
JP3_意見文	57	2.11	-4.05	-0.85	-1.06

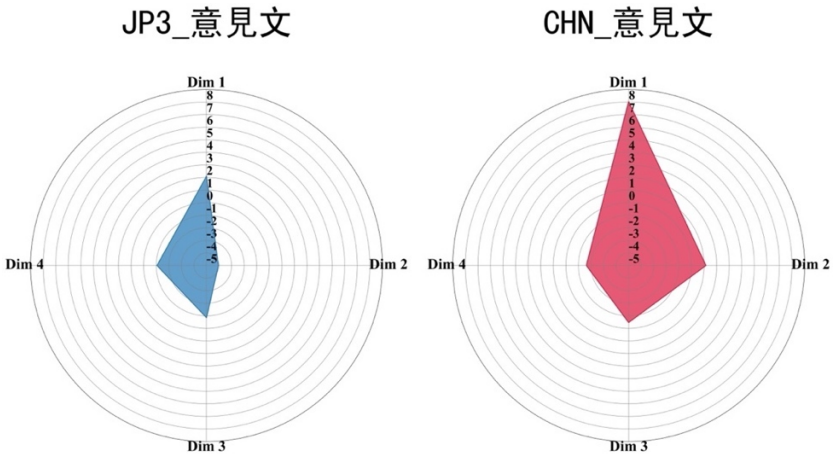

図 6.10　「意見文」：母語話者作文と学習者作文の比較（radar chart）

表6.18　「意見文」に関する各次元スコアの多重比較

		次元 1		次元 2		次元 3		次元 4	
Group1	**Group2**	**Diff**	**Sig.**	**Diff**	**Sig.**	**Diff**	**Sig.**	**Diff**	**Sig.**
CHN_意見文	JP3_意見文	5.922	0.001	5.208	0.001	0.382	0.069	-0.579	0.066

注：「Diff」は、Group1 と Group2 の平均値の差であり、「Sig.」は *p*-value である。

母語話者と学習者が書いた「意見文」作文ジャンルは各次元での平均値の差が統計的に有意であるかどうかを確認した結果（表 6.18）、次元 3「副詞の修飾」と次元 4「動

作の描写」について、学習者（JP2）と母語話者（CHN）の間に有意差は認められないものの（次元 3：$p = 0.069 > 0.05$; 次元 4：$p = 0.066 > 0.05$）、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」は、CHN > JP2 の順で 、学習者（JP2）と母語話者（CHN）の間（次元 1：$p = 0.001 < 0.05$; 次元 2：$p = 0.001 < 0.05$）に有意差があることを確認した。意見文のような作文を書く際、学習者（JP2）の方が産出語彙の豊富さと状態状況表現の適切さが低い傾向があることが分かった。

(6) まとめ

表 6.19 は、同一ジャンルにおける母語話者作文と学習者作文の多重比較の結果をまとめたものである。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、学習者が母語話者より、作文における語彙が主に低難度語彙に限定しており、産出語彙の豊富さと状態状況表現の適切さが低く、副詞の修飾が少ない傾向があることが明らかになった。

表 6.19 同一作文ジャンルにおける母語話者作文と学習者作文：多重比較の結果のまとめ

	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
出来事の叙述	CHN > JP3 > JP2	CHN > JP2 > JP3	CHN > JP3 ≈ JP2	CHN ≈ JP2 > JP3
人物描写	CHN > JP3 > JP2	CHN > JP3 ≈ JP2	CHN > JP3 > JP2	CHN > JP3 ≈ JP2
手紙	CHN > JP2	CHN > JP2	CHN > JP2	CHN ≈ JP2
日記	CHN > JP2	CHN > JP2	CHN > JP2	CHN ≈ JP2
意見文	CHN > JP3	CHN > JP3	CHN ≈ JP3	CHN ≈ JP3

注：「> 」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

6.3 本章のまとめ

本章では、中国語作文における言語変異の次元を反映する 4 つの共起パターンを主な次元として捉え、各次元を構成する有効な言語項目の伝達機能を解釈し、ジャンル間の類似点と相違点を考察した。

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

本章では、第 6 章に示した次元１～4 を構成する有効な言語項目に基づき、同一言語項目における学習者作文と母語話者作文ジャンル間の比較を行う。各言語項目について、サブコーパス別の相対頻度は箱ひげ図（Boxplot）[75]で平均値の昇順に表示される。

7.1 次元１の言語項目

表 7.1 は、次元１を構成する 7 つの言語項目の相対頻度の平均値を示している。

表 7.1　次元１に属する言語項目の相対頻度（平均値）

Sub corpus	高難度語彙	語彙多様性	非 HSK 語彙	平均文長	韻律双音節語	一人称単数代名詞“我”	低難度語彙
CHN_人物描写	53.73	11.06	184.99	28.16	6.96	31.91	348.44
CHN_出来事の叙述	61.82	11.40	201.38	29.61	9.39	29.40	317.64
CHN_意見文	85.16	12.26	210.25	35.38	17.61	7.77	240.37
CHN_手紙	61.03	11.57	185.33	30.02	10.93	29.58	334.36
CHN_日記	62.28	11.14	197.10	30.15	10.19	22.69	315.06
JP2_人物描写	37.63	8.46	149.05	17.85	8.01	56.01	367.32
JP2_出来事の叙述	24.77	8.50	155.26	17.88	7.43	60.78	387.57
JP2_手紙	31.26	9.40	145.96	18.53	7.98	45.79	359.71
JP2_日記	36.54	8.87	160.08	20.37	8.00	39.09	354.44
JP3_人物描写	43.98	9.30	139.28	21.89	9.63	40.76	372.38
JP3_出来事の叙述	45.74	10.18	164.90	22.43	12.02	29.07	318.86
JP3_意見文	64.12	9.41	162.66	25.09	17.18	19.20	279.40

[75] 徐(2022)：「箱ひげ図では、データのばらつきを分かりやすく表現することができる。点(dots)は個別の作文の統計量を表し、横軸にサブコーパス名、縦軸に比較項目の統計量が配置されている。箱ひげ図により、比較項目におけるサブコーパスごとの違いが見られる」。

（1）平均文長と語彙多様性

「平均文長」はセンテンスの長さを表す指標であり、「語彙多様性」はテキストにおける産出した語彙の豊富さを量的に測る指標である。語彙多様性と平均文長の Boxplot（図 7.1~図 7.2）から明らかなように、学習者の方が母語話者より全体的に作文の平均文長が低く、語彙多様性が低い傾向が見られる。

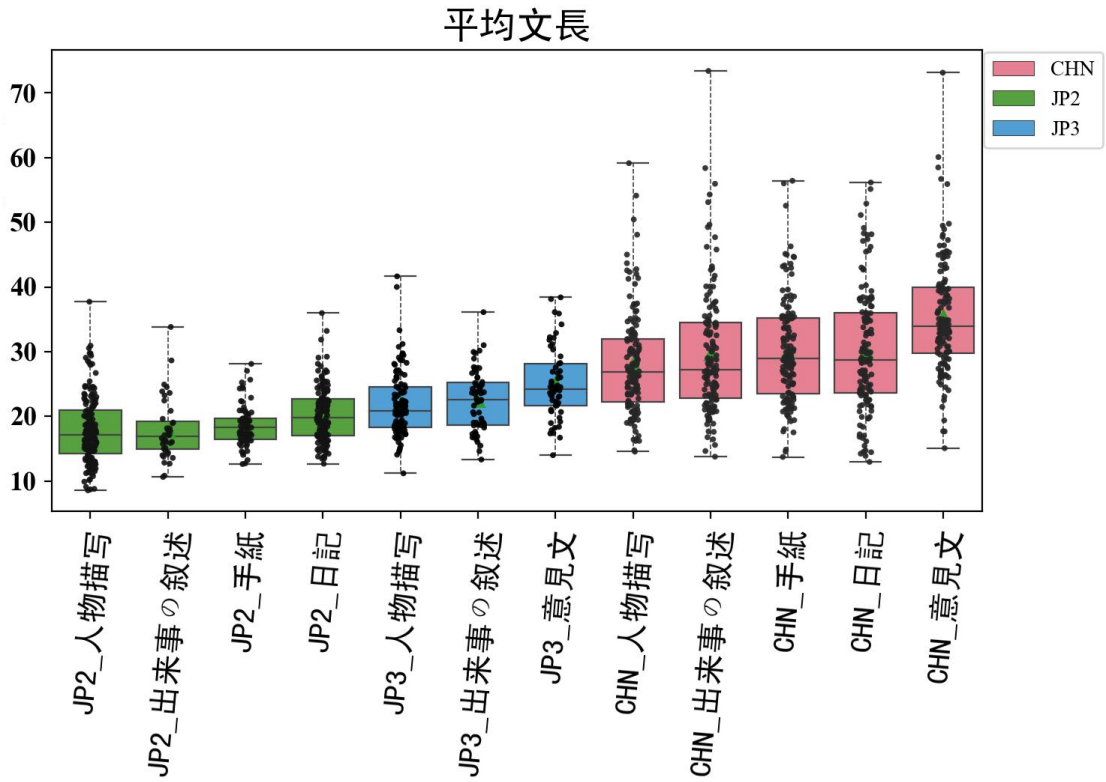

図 7.1　平均文長の Boxplot

母語話者作文ジャンル間について多重比較を行った結果、平均文長も語彙多様性も、「意見文」と他の 4 つのジャンル間に有意差があり（$p < 0.05$）、「日記」・「出来事の叙述」・「手紙」・「人物描写」という 4 つのジャンル間にはお互いに有意差がない（$p > 0.05$）ことが確認された。ほかのジャンルに比べ、「意見文」は平均文長が著しく長く、語彙多様性が著しく高いことが分かった。平均文長も語彙多様性も高い順に、「意見文＞日記 ≈ 手紙 ≈ 出来事の叙述 ≈ 人物描写」となっている。

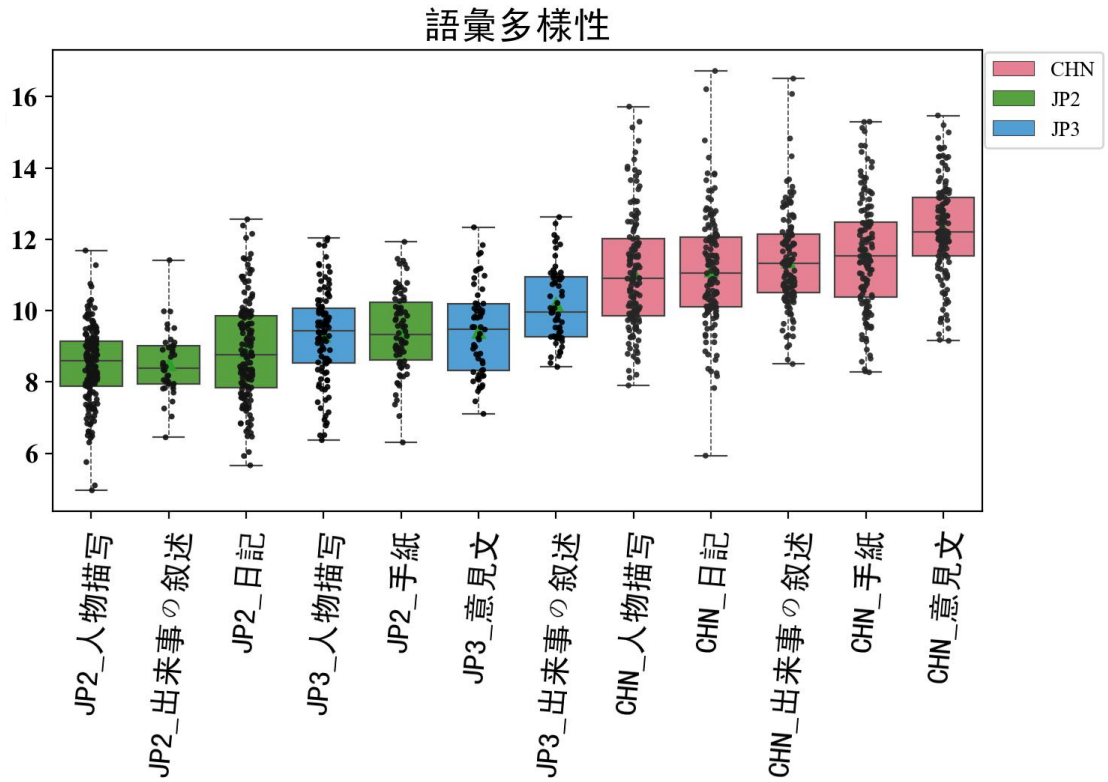

図 7.2　語彙多様性の Boxplot

平均文長と語彙多様性について、同一作文ジャンルにおける母語話者と学習者作文を比較するために、多重比較も行った。その結果（表 7.2）、平均文長も語彙多様性も、学習者（JP2・JP3）と母語話者（CHN）の間に有意差があり（$p < 0.05$）、学習者の方が作文における平均文長と語彙多様性が明らかに低いことを確認した。

表 7.2　母語話者と学習者作文における平均文長と語彙多様性の多重比較

	平均文長	語彙多様性
出来事の叙述	CHN > JP3 ≈ JP2	CHN > JP2 > JP3
人物描写	CHN > JP3 > JP2	CHN > JP3 > JP2
手紙	CHN > JP2	CHN > JP2
日記	CHN > JP2	CHN > JP2
意見文	CHN > JP3	CHN > JP3

注：「>」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

例文（18）～（20）では、CHN・JP3・JP2 サブコーパスから抽出された人物描写の作文の抜粋を示している。下記の例文に示すように、母語話者に比べ、学習者の作文、特に JP2（中級学習者）の作文は、センテンスの長さが短く、短いセンテンス（短文）が積み重なっており、語彙の繰り返しが多い。

例文：

(18) 今天 ， 我 要 偷偷 地 把 一个 秘密 告诉 你们 ， 我 有 三 个 奶奶 ， 相信 大 家 听 了 我 的 话 之后 会 觉得 很 奇怪 ， 有些 人 会 说 ： “ 怎么 可能 ， 一个 人 最 多 只 有 两 个 奶奶 。 ” 不过 我 这 第三 个 奶奶 是 我 外婆 的 邻居 。

（CHN_人物描写_049, Dimension 1 score: - 0.95, 平均文長: 38.56, 語彙多様性: 11.71）

(19) 我 和 他 是 15 年 前 就 认识 的 朋友 。 第一 次 我们 见面 的 时候 是 在 小 学校 的 时候 ， 一 年级 的 学生 。 一方面 当时 的 我 是 容易 怕生 的 人 ， 没有 多 朋友 。 另一方面 ， 他 不 怯生 ， 有 很多 朋友 。

（JP3_人物描写_064, Dimension 1 score: - 4.68, 平均文長: 21.46, 語彙多様性: 9.74）

(20) 她 是 大阪 人 。 她 家 有 四 口 人 。 爸爸 ， 妈妈 ， 弟弟 ， 和 她 。 她 也 有 一 条 泥鳅 ， 它 很 可爱 ！ 她 在 补 学 班 打工 。 她 教 英语 ， 历史 ， 数学 等等 。

（JP2_人物描写_020, Dimension 1 score: - 4.91, 平均文長: 14.75, 語彙多様性: 8.91）

---

(2) 非 HSK 語彙・高難度語彙・低難度語彙

まず、非 HSK 語彙について、図 7.3 の Boxplot から明らかなように、学習者に比べて、母語話者の作文における非 HSK 語彙の出現頻度が全体的に高い傾向が見られる。高難度語彙に関しては、「JP3_意見文」以外に、学習者の方が母語話者より全体的に高難度語彙の使用頻度が低い傾向が見られる（図 7.4）。一方、図 7.5 に示すように、「JP3_人物描写」以外に、JP3 と母語話者のジャンルに比べて、JP2 作文ジャンルの方が全体的に低難度語彙の使用頻度が高い。

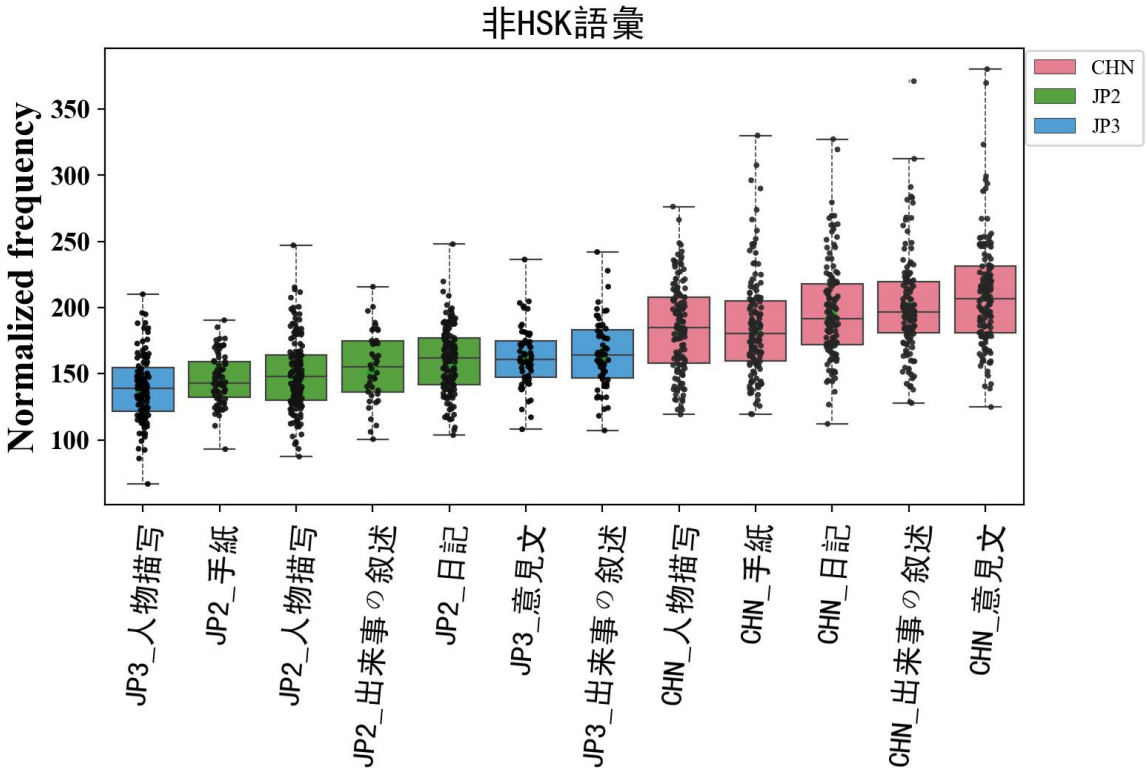

図 7.3　非 HSK 語彙の Boxplot

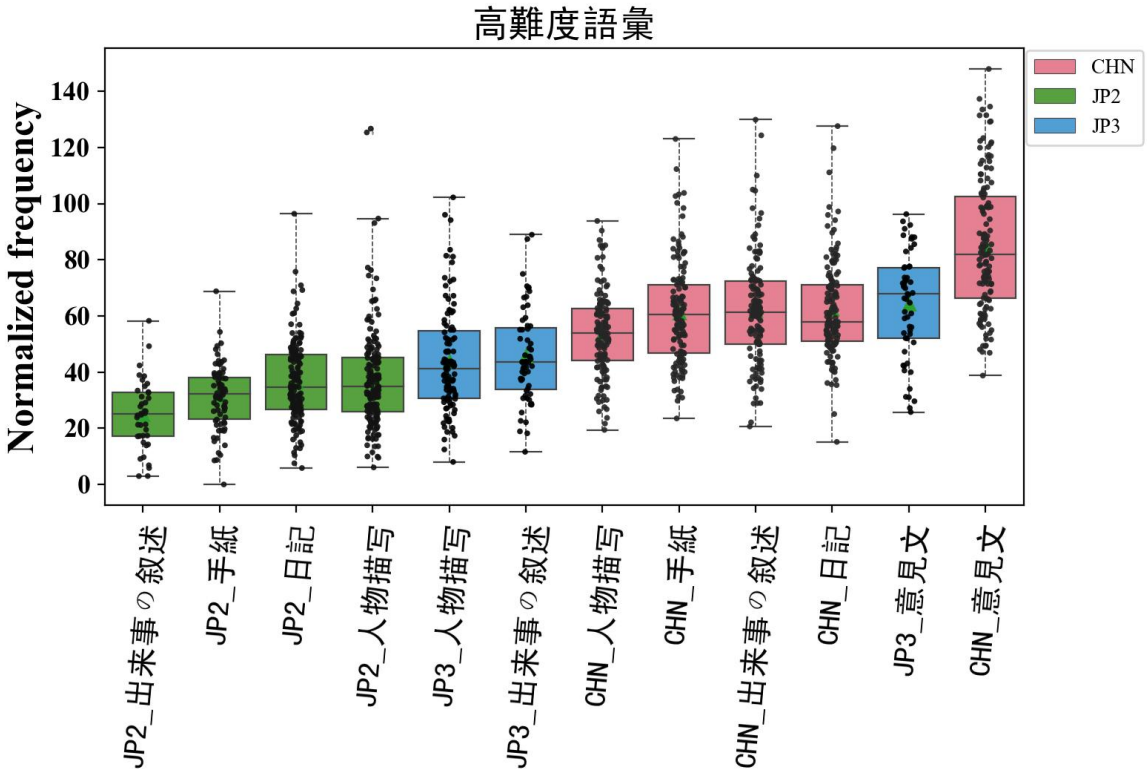

図 7.4　高難度語彙の Boxplot

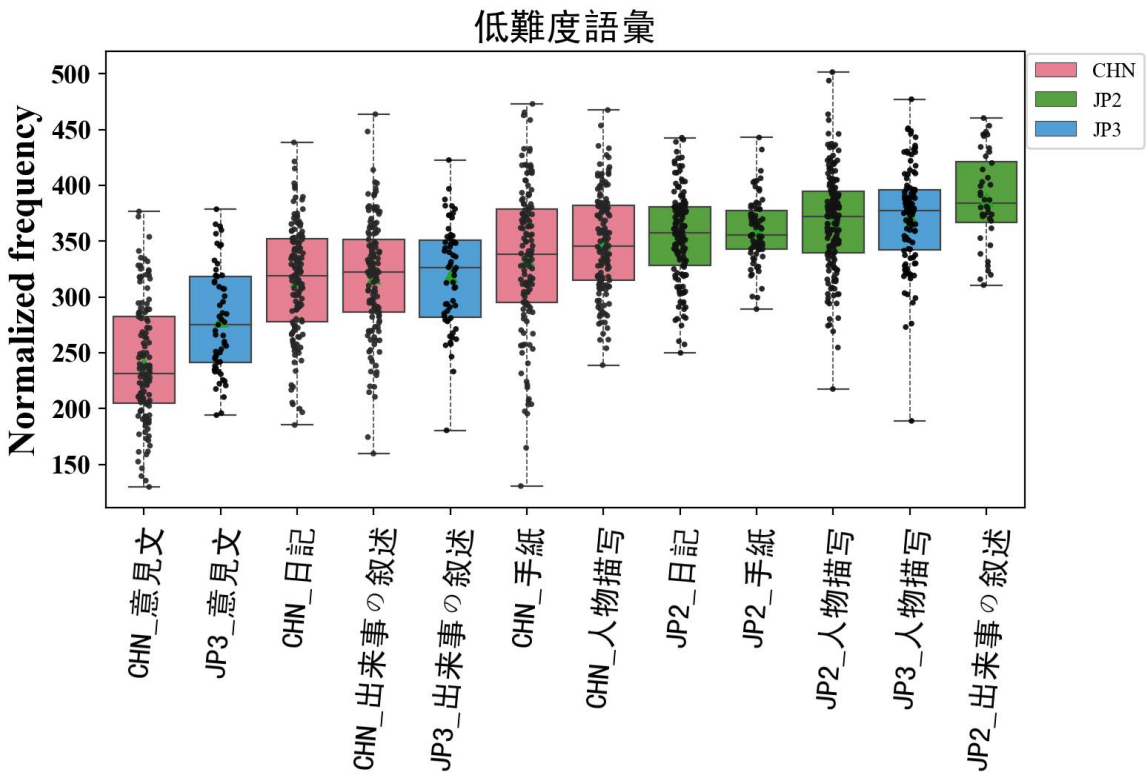

図 7.5　低難度語彙の Boxplot

これら 3 つの言語項目について、同一作文ジャンルにおける母語話者作文と学習者作文を比較するために、多重比較を行った。その結果（表 7.3）、非 HSK 語彙も高難度語彙も、学習者（JP2・JP3）と母語話者（CHN）の間に有意差があり（$p < 0.05$）、学習者の方が作文における非 HSK 語彙と高難度語彙の使用頻度が明らかに低いことを確認した。一方、低難度語彙の使用頻度は、学習者（JP2 / JP3）と母語話者（CHN）の間に有意差があり（$p < 0.05$）、学習者の方が作文における低難度語彙の使用頻度が著しく高いことが確認された。

表 7.3　母語話者と学習者作文における非 HSK 語彙・高難度語彙・低難度語彙の多重比較

	非 HSK 語彙	高難度語彙	低難度語彙
出来事の叙述	CHN > JP3 ≈ JP2	CHN > JP3 > JP2	JP2 > JP3 ≈ CHN
人物描写	CHN > JP3 ≈ JP2	CHN > JP3 ≈ JP2	JP3 ≈ JP2 > CHN
手紙	CHN > JP2	CHN > JP2	JP2 > CHN
日記	CHN > JP2	CHN > JP2	JP2 > CHN
意見文	CHN > JP3	CHN > JP3	JP3 > CHN

注：「>」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

(3) 韻律双音節語

第 2 章で説明したように、「韻律双音節語」は話し言葉では稀で、現代中国語の書き言葉にのみ現れる（馮・王・黄 2008）言語項目である。あるテキストには韻律双音節語が高頻度で出現する場合、そのテキストはより正式的・書面的であるものだといえる。図 7.6 に示したように、韻律双音節の頻度について、意見文は他の作文ジャンルより全体的に高い傾向が見られる。

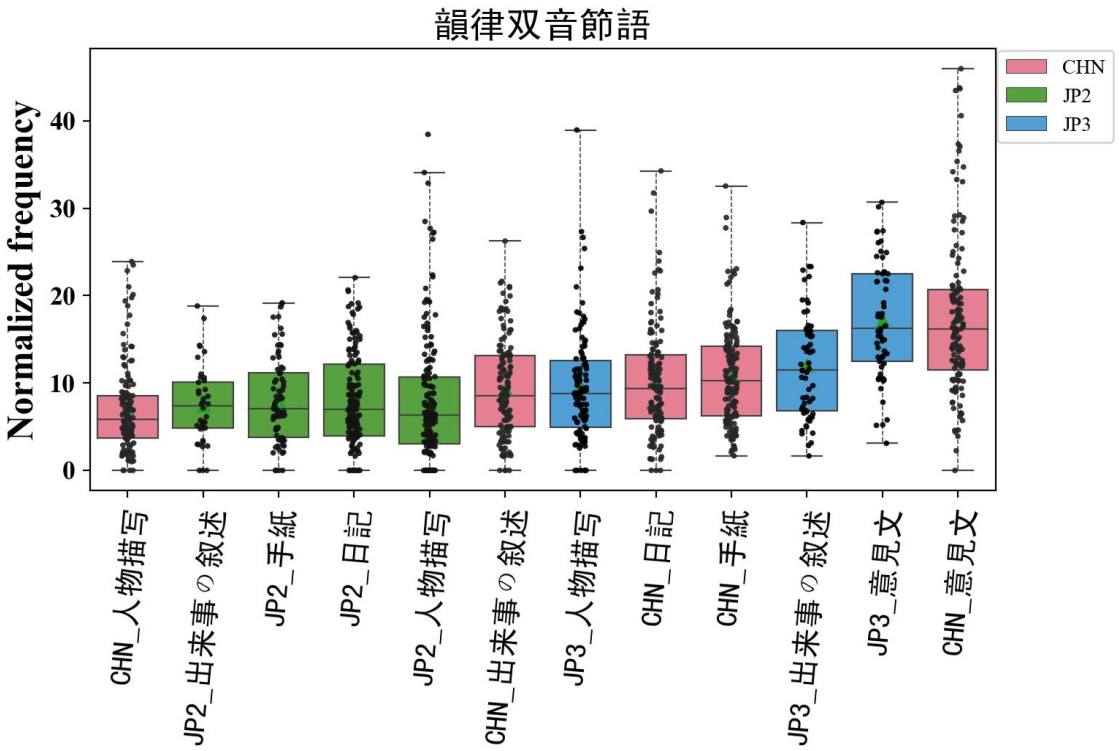

図 7.6　韻律双音節語の Boxplot

同一ジャンルにおける母語話者作文と学習者作文で使用された韻律双音節語について、多重比較を行った。その結果（表 7.4）、同一作文ジャンルにおける韻律双音節語の使用頻度は、学習者（JP2 / JP3）と母語話者（CHN）の間に有意差が認められない（$p > 0.05$）ものの、学習者と母語話者の意見文は他のジャンルより、韻律双音節語の使用頻度が著しく高い（$p < 0.05$）ことが明らかになった。

表 7.4　母語話者と学習者作文における韻律双音節語の多重比較

	韻律双音節語
出来事の叙述	CHN ≈ JP3 ≈ JP2
人物描写	CHN ≈ JP3 ≈ JP2
手紙	CHN ≈ JP2
日記	CHN ≈ JP2
意見文	CHN ≈ JP3

注：「>」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

出来事の叙述や人物描写の作文では、出来事や人物の特徴を際立たせるために、直接話法で聞いた言葉を直接引用することが多く、話し言葉の表現（例、例文 21 における“兔崽子”：ろくでなしめ）もよく見受けられる。一方、意見文は、推論と論証に重点を置き、ある事実や現象に対して、賛成とか反対とか、自分の意見をきちんとした理由で説明し、より論理的な作文であるため、書面性の高い韻律双音節（例、 例文 22 におけるの“成为”、“建立”、“如同”など）の使用が多いと推察される。

例文：

(21) 我 的 姥爷 也 就 是 一个 普普通通 的 农民 ， 黝黑 的 皮肤 ， “ 褶皱 ” 的 脸 。 我 觉得 我 很 少 能够 亲近 他 。小时候 ， 姥爷 对 我 非常 严厉 ， 我 做 什么 他 都 会 连 说 带 骂 地 来 两 句 。 “ 你 个 兔崽子 …… ” 这 是 他 对 我 说 的 最 多 的 一 句 话 了 。

（CHN_人物描写_118, Dimension 1 score: - 0.33, 韻律双音節語: 8.24）

(22) 新型 冠状 肺炎 已 在 全球 蔓延 ， 这 是否 又 是 自然 在 向 人类 社会 发出 警告 呢 ？ 曾 高呼 “ 不要 口罩 要 自由 ” 的 意大利 已 成为 欧洲 的 疫情 重灾区 ！ 经过 漫长 的 历史 岁月 ， 人类 建立 一 套 社会 法则 ， 构建 起 一 座 庞大 的 精神 大厦 ， 可是 ， 在 自然 法则 面前 ， 人类 如同 沧海 一 粟 ， 渺小 而 不堪一击 。 (CHN_意見文_039, Dimension 1 score: 12.21, 韻律双音節語: 19.55)

---

(4) 一人称単数代名詞“我”

図 7.7 に示すように、「JP3_人物描写」以外に、JP2 は JP3 と母語話者より、作文における一人称単数代名詞“我”(私)の使用頻度が全体的に高い傾向が見られる。

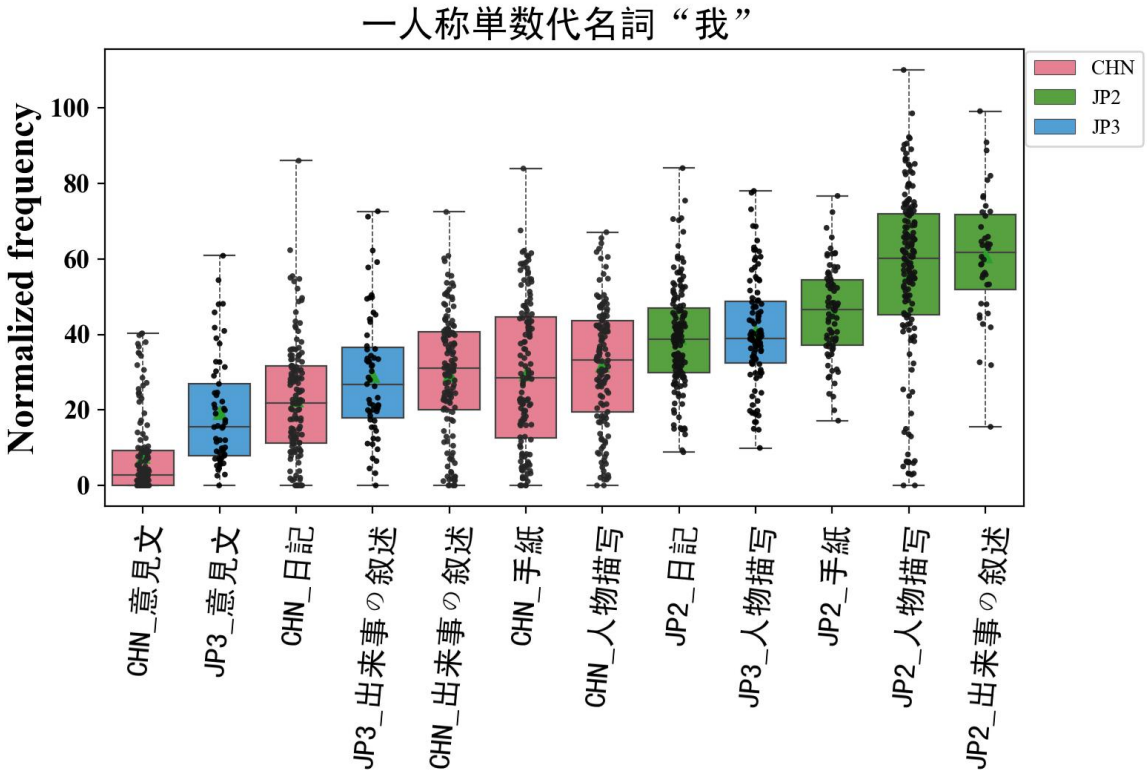

図 7.7　一人称単数代名詞“我”の Boxplot

母語話者作文ジャンル間の多重比較を行った結果、一人称単数代名詞“我”の使用頻度の高い順は、「人物描写 ≈ 手紙 ≈ 出来事の叙述 >日記 > 意見文」となっている。つまり、「人物描写・手紙・出来事の叙述」の 3 つのジャンルはお互いに有意差が認められ

ない（$p > 0.05$）ものの、これら 3 つのジャンルはそれぞれ、「日記 」や「意見文」より著しく高い（$p < 0.05$）ことが確認された。また、「意見文」より「日記 」の方が、一人称単数代名詞“我”の使用頻度が著しく高い（$p < 0.05$）ことも明らかになった。

同一ジャンルにおける母語話者と学習者作文で使用された一人称単数代名詞“我”（私）の使用頻度について、多重比較を行った結果（表 7.5）、学習者（JP2 / JP3）と母語話者（CHN）の間に有意差があり（$p < 0.05$）、学習者は母語話者より、各作文ジャンルにおける一人称単数代名詞“我”の使用頻度が著しく高いことを確認した。

表 7.5　母語話者と学習者作文における一人称単数代名詞“我”の多重比較

	一人称単数代名詞“我”
出来事の叙述	JP2 > JP3 ≈ CHN
人物描写	JP2 > JP3 > CHN
手紙	JP2 > CHN
日記	JP2 > CHN
意見文	JP3 > CHN

注：「> 」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

7.2 次元 2 の言語項目

次元 2 に属する言語項目の相対頻度の平均値は表 7.6 に示している。図 7.8~図 7.12 は、次元 2 における各言語項目の相対頻度の Boxplot（箱ひげ図）である。

表 7.6　次元 2 の言語項目の相対頻度（平均値）

Sub corpus	時態助詞 “着”	ぼかし 言葉	程度を 表す副詞	接続詞	単語の 平均長さ
CHN_人物描写	7.55	18.23	7.73	14.05	1.42
CHN_出来事の叙述	7.63	18.93	7.11	14.67	1.45
CHN_意見文	3.50	17.15	6.30	18.96	1.56
CHN_手紙	4.77	19.90	8.26	16.39	1.47
CHN_日記	5.79	22.06	9.06	15.31	1.48
JP2_人物描写	1.11	32.35	18.43	26.97	1.53
JP2_出来事の叙述	1.90	28.26	15.46	24.85	1.50
JP2_手紙	0.60	35.78	18.21	24.27	1.57
JP2_日記	1.05	32.74	19.86	26.02	1.56
JP3_人物描写	1.41	32.95	16.49	27.12	1.50
JP3_出来事の叙述	0.62	31.84	16.09	27.04	1.59
JP3_意見文	0.20	38.31	12.80	31.57	1.61

（1）時態助詞“着”

まず、「時態助詞“着”」は、主に「Verb +“着”」の使い方で、動詞の後ろに置き、動作や状態の持続を表す表現である。図 7.8 から見られるように、JP2 と JP3 は、母語話者に比べ、作文における「時態助詞“着”」の使用頻度が全体的に低い傾向が見られる。One-way ANOVA 分析を行った多重比較の結果、母語話者の作文ジャンルにおける「時態助詞“着”」の使用頻度に関しては、「出来事の叙述 ≈ 人物描写 > 日記 ≈ 手紙 > 意見文」[76]であり、母語話者の「出来事の叙述」と「人物描写」という 2 つの作文ジャンルには、「時態助詞“着”」の使用頻度は一番顕著であることを確認した。

[76] 「＞」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

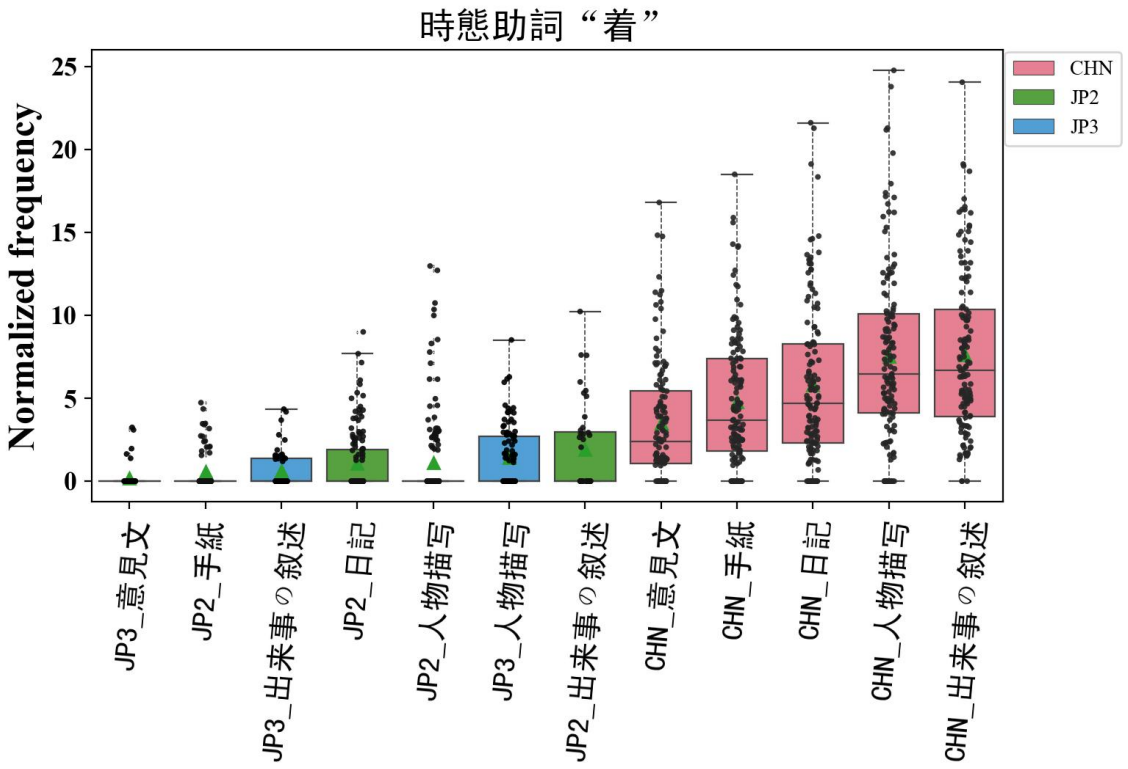

図 7.8　時態助詞“着”の Boxplot

同一作文ジャンルにおける母語話者と学習者作文で使用された時態助詞“着”について、多重比較も行った。その結果（表 7.7）、同一作文ジャンルにおける「時態助詞“着”」の使用頻度は、学習者（JP2・JP3）と母語話者（CHN）の間に有意差があり（$p < 0.05$）、学習者は母語話者より、各作文ジャンルにおける「時態助詞“着”」の使用頻度が著しく低いことを確認した。

表 7.7　母語話者と学習者作文における時態助詞“着”の多重比較

	時態助詞“着”
出来事の叙述	CHN > JP3 ≈ JP2
人物描写	CHN > JP3 ≈ JP2
手紙	CHN > JP2
日記	CHN > JP2
意見文	CHN > JP3

注：「>」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

(2) 単語の平均長さ

本研究で特定された単語の平均長さ（Average word length）とは、テキストにおける単語の平均的な長さを文字数で表す指標である。図 7.9 に示したように、「CHN_意見文」以外に、学習者は母語話者より作文における単語の平均長さが高い傾向が見られる。

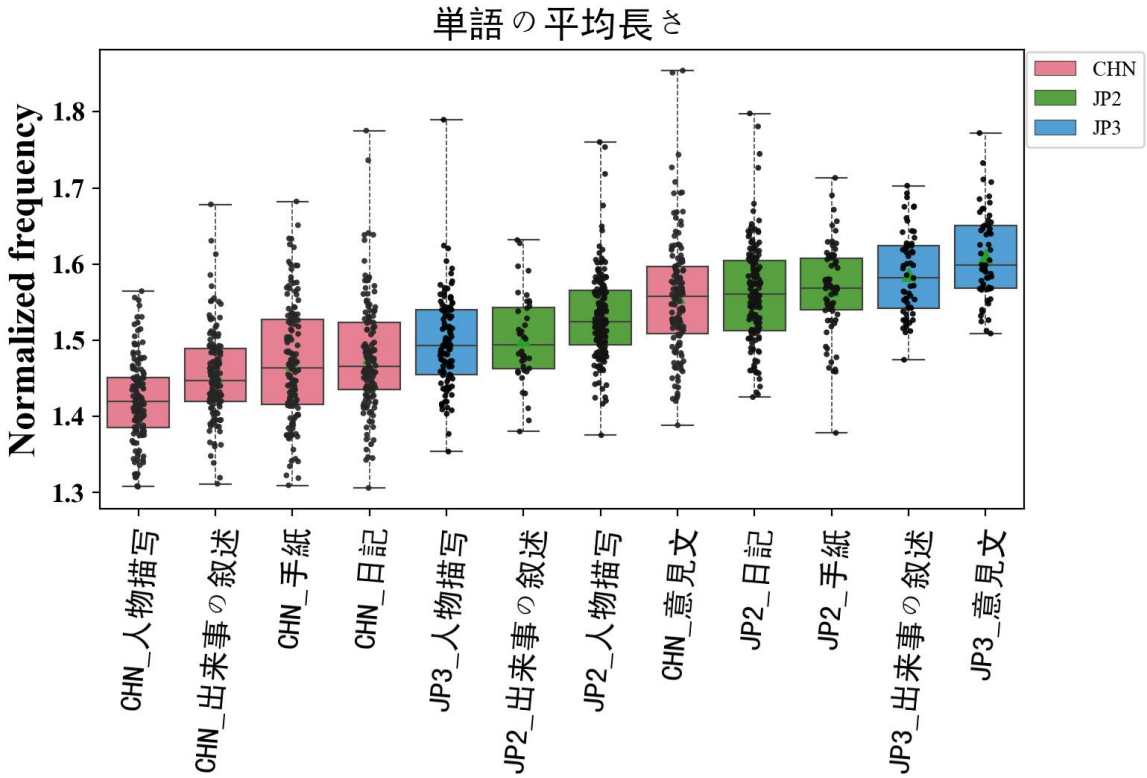

図 7.9　単語の平均長さの Boxplot

同一作文ジャンルにおける母語話者と学習者作文の単語の平均長さについて、多重比較を行った結果（表 7.8）、学習者（JP2 / JP3）と母語話者（CHN）の間に有意差があり（$p < 0.05$）、学習者は母語話者より、各作文ジャンルにおける単語の平均長さが著しく長いことを確認した。

表 7.8　母語話者と学習者作文における単語の平均長さの多重比較

	単語の平均長さ
出来事の叙述	JP3 > JP2 > CHN
人物描写	JP2 > JP3 > CHN
手紙	JP2 > CHN
日記	JP2 > CHN
意見文	JP3 > CHN

注：「> 」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

その原因は、学習者の作文における日本語の人名・団体名・地名が多く、これらの単語の長さは、中国語母語話者の作文における単語よりも一般的に長いからであると考えられる。例えば、例文（23）において、「水溜りボンド」は日本の 2 人組 YouTuber の名前であり、この単語の長さは 6 字数である。

例文：

（23）昨天 晚上 ， 我 偶然 在 YouTube 上 发现 了 一个 视频 。 这 是 我 喜欢 的 YouTuber 公开 的 视频 。 他们 是 【 水溜りボンド 】 ， 是 很 有名 的 日本 人 的 YouTuber 。 这个 视频 表达 了 【 水溜りボンド 】 的 一个 成员 「 トミー 」 一个 星期 中 在 东京 的 森林 里 体验 露营 生活 。

（JP3_出来事の叙述_029, Dimension 2 score: -1.07, 単語の平均長さ: 1.53）

---

(3) ぼかし言葉・程度を表す副詞

本研究で特定された「程度を表す副詞」という言語項目は「ぼかし言葉」の一部である。図 7.10 と図 7.11 は、各サブコーパスにおける「ぼかし言葉」と「程度を表す副詞」の使用頻度の箱ひげ図である。図 7.10〜図 7.11 から明らかなように、作文を書く際、学習者の方が母語話者より全体的に「ぼかし言葉」、「程度を表す副詞」の使用頻度が高い傾向が見られる。

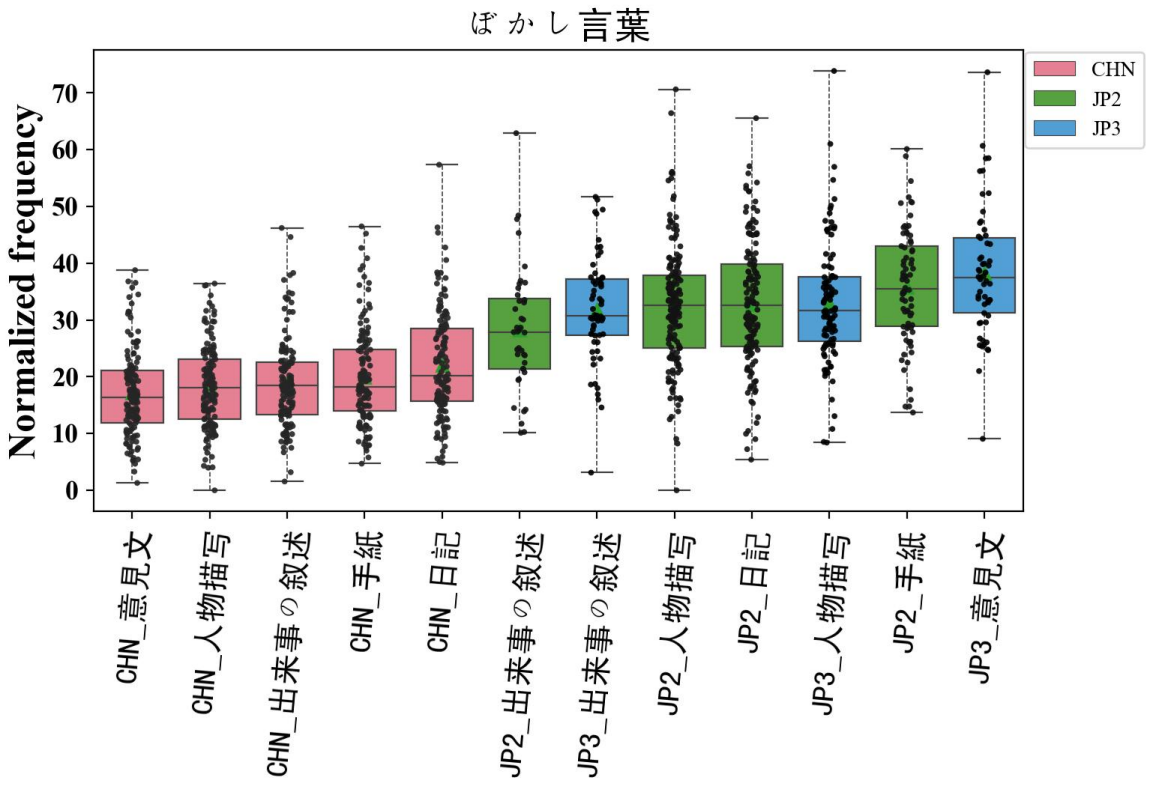

図 7.10　ぼかし言葉の Boxplot

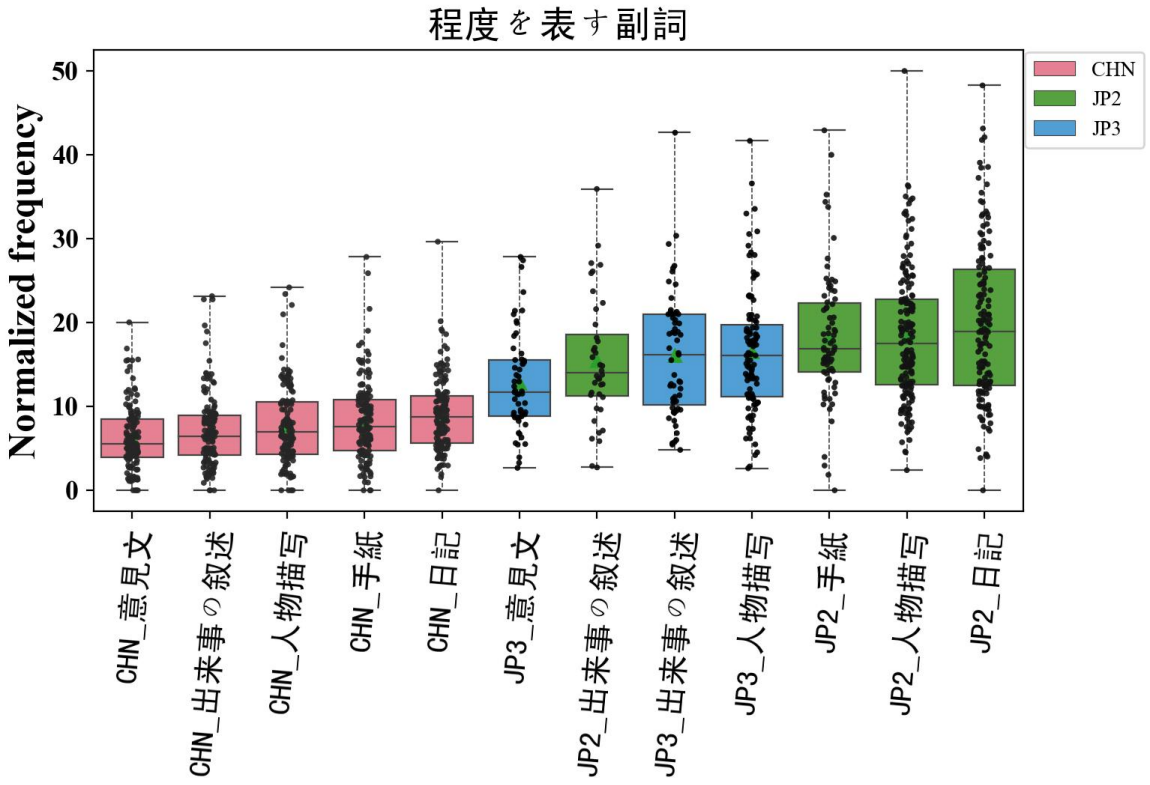

図 7.11　程度を表す副詞の Boxplot

One-way ANOVA 分析を行った多重比較の結果（表 7.9）、同一作文ジャンルにおける「ぼかし言葉」も「程度を表す副詞」も、学習者（JP2 ・ JP3）と母語話者（CHN）の間に有意差があり（$p < 0.05$）、学習者は母語話者より、各作文ジャンルにおける「ぼかし言葉」と「程度を表す副詞」の使用頻度が著しく高いことを確認した。

表 7.9　母語話者と学習者作文における「ぼかし言葉」と「程度を表す副詞」の多重比較

	ぼかし言葉	程度を表す副詞
出来事の叙述	JP3 ≈ JP2 > CHN	JP3 ≈ JP2 > CHN
人物描写	JP2 ≈ JP3 > CHN	JP2 ≈ JP3 > CHN
手紙	JP2 > CHN	JP2 > CHN
日記	JP2 > CHN	JP2 > CHN
意見文	JP3 > CHN	JP3 > CHN

注：「>」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

下記の例文（24）～例文（27）は、学習者と母語話者が書いた意見文及び日記の抜粋である。例文（27）～例文（28）を合わせてみると、学習者の作文において、ぼかし表現に属する数量詞（例、“很多”：たくさん）、形容詞（例、“多”：多い；“很好”：素晴らしい；“好”：よい）、副詞（例、“可能”：かもしれない）、程度を表す副詞（例、“*很*”：とても；“*非常*”：非常に）などは多く使用されていることが分かった。

例文：

(24) 我 不 赞成 在 疫情 中 不 戴 口罩 。 我 支持 戴 口罩 的 理由 有 一个 ： 因为 如果 现在 不 戴 口罩 就 直接 出门 的 话 ， *很* 有 可能 会 让 无辜 的 人 感染 新冠 肺炎 。 也就是说 ， 戴 口罩 不仅 能 让 我们 预防 感染 ， 还 能够 预防 把 病毒 传染 给 别人 。 很多 西方人 可能 都 以为 除非 太阳 从 西边 出来 ， 不然 他们 不 会 感染 新冠 肺炎 ， 所以 不 戴 口罩 也 没关系 。 不过 ， 这个 想法 可以 说 是 *非常* 地 我行我素 。 我们 不 应该 只顾 自己 ， 也 应该 要 在

乎 会 不 会 影响 到 别人 才 对 。 如果 他们 像 这样 继续 自私 下去 的 话 ， 会 *很* 难 让 感染 人数 减少 。 他们 应该 要 重新 想 一 想 如果 因为 自己 不 适当 的 行为 而 让 别人 感染 的 话 ， 他们 负 不 负 得 起 责任 。 即便 是 在 疫情 *特别* 严重 的 情况 之下 ， 为什么 有 这么 多 包含 美国人 在内 的 西方 人 不 戴 口罩 呢 ？ 在 一个 由 中国 记者 在 伦敦 所 做 的 采访 中 ， 很多 英国 人 说 那 是 因为 在 英国 如果 戴 口罩 出门 的 话 ， 会 *很* 显眼 。

（JP3_意見文_009, Dimension 2 score: -3.89, ぼかし言葉: 47.08, 程度を表す副詞: 14.61）

(25) 二零二零年 一月 中旬 ， 当 大家 都 为 即将 到来 的 春节 忙碌 准备 时 ， 新型 冠状 肺炎 病毒 悄无声息 地 爆发 了 。 而 网络 授课 也 成为 一 种 新 的 教学 模式 出现 在 我们 面前 。 这 一 措施 刚 开始 实行 ， 便 引起 广大 学生 ， 家 长 和 教学 职工 的 热 议 ， 有人 对 网课 app 给 出 “ 一 星 好评 ” ， 也 有 人 认为 这 是 一 次 赶超 他人 的 好 机会 。 在 我 看来 ， 网络 授课 期间 对 学生 *最* 大 的 考验 就 是 自觉性 和 自律性 。 对 老师 和 家长 的 考验 便是 责 任感 。 首先 网络 授课 是 把 双刃剑 ， 有利 也 有 弊 。 各种 授课 app 的 出 现 使得 教育 方式 变得 *更加* 灵活 ， 全面 的 功能 也 让 老师 的 监督 工作 *更 加* 便利 ， 让 同学 们 的 学习 效率 有所 提高 。

（CHN_意見文_041, Dimension 2 score: -2.45, ぼかし言葉: 19.47, 程度を表す副詞: 6.95）

(26) 2020 年 1 月 1 日 晴天

这 一 天 ， 我 去 了 奶奶 的 家 ， 开 了 新年会 。 我 吃 了 年节 菜 ， 寿司 和 螃蟹 。 那种 菜 在 我 的 家庭 特别 的 料理 ， *非常* 好吃 ！ 吃饭 之后 ， 我 和 妈妈 、 奶奶 、 爷爷 和 我 妈妈 的 弟弟 一起 打 扑克 比赛 。 我们 打 了 各种各样 的 游戏 ， 过 得 *很* 开心 ！ 那 时候 ， 奶奶 对 在 游戏 里 第一 名 的 人 准备 点心 等 很多 奖品 。 我 也 得 了 很多 彩 了 。 这 是 我 家人 的 过 新年 的 方法 ， 我 的 信念 第一 的 期待 ！

2020 年 1 月 13 日 晴天

这 一 天 是 我 的 成人 仪式 。 为了 穿 上 宽 幅 ， 我 5 点 起床 。 在 成人 仪式 的 会场 ， 和 很多 中学 同学 拍照 、 聊天 、 过 得 *非常* 开心 。 朋友 们 的 衣服 *很* 漂亮 ， 化妆 很好 ！ 成人 仪式 以后 ， 我 和 妈妈 ， 老爷 ， 姥姥 一起 去 吃饭 。 那个 餐厅 是 我们 家 常 去 的 ， 吃 了 美味 的 牛排 ！ 那天 ，

我 也 开车 送 姥姥 车站 ， 那时 是 好久 的 开车 ， 所以 很 紧张 。 不过 ，
我 开车 开 得 很好 。 这 一 天 ， 我 不 会 忘 了 ， 好 回忆 ！

（JP2_日記_022, Dimension 2 score: -4.19, ぼかし言葉: 38.29, 程度を表す副詞: 35.49）

（27）2020 年 7 月 5 日 多云

早上 ， 我 还 沉浸 在 睡梦 之 时 ， 就 隐约 听到 了 说话 的 声音 。 我 不 情 愿 睁 开 眼睛 ， 听 到 那 声音 是 表弟 的 房间 传来 的 ， 想不到 小 表弟 已 经 起床 ， 正 和 爷爷 奶奶 聊天 ， 我 三下五除二 穿 好 衣服 来到 表弟 的 房 间 ， 和 表弟 一起 玩 了 一会儿 扑克牌 。 今天 的 学习 任务 就 正式 开始 了 ！ 我 先 教 表弟 背 两 首 唐诗 ， 这 两 首 唐诗 分别 是 张说 的 《 送梁六自洞庭 山 》 和 张敬忠 的 《 边词 》 。 表弟 一 见 这 两 首 诗 都 是 七 言 绝句 便 一 脸 不满 ， 开始 了 大 罢工 。 小样 ， 还 想 跟 我 斗 。 我 心 生 一 计 便 对 表弟 说 ： “ 昨天 你 背 的 的 两 首 唐诗 都 是 五言 绝句 ， 总共 是 八 句 。 八 乘 五 等于 “ 八十 ” ， 总共 是 “ 八十 ” 个 字 。 今天 你 背 的 两 首 唐诗 是 七言 绝句 ， 总共 也 是 八 句 。 八 乘 七 等于 五十六 ， 你 比 昨天 还 少 背 了 二十四 个 字 呢 。 ” 幸好 表弟 才 上 一 年级 ， 还 不 会 乘法 ， 竟 相信 了 我 的 “ 无稽之谈 ” 。 事 成 之后 我 一 脸 奸笑 ， 不料 却 被 表弟 发觉 ， 他 又 开始 了 罢工 。 我 再次 对 他 连 唬 带 骗 ， 把 他 给 糊弄 了 过去 。

（CHN_日記_088, Dimension 2 score: 3.96, ぼかし言葉: 17.72, 程度を表す副詞: 3.80）

（4）接続詞

図 7.12 に示したように、学習者の方が母語話者より、全体的に作文における「接続詞」の使用頻度が高い傾向も見られる。同一作文ジャンルにおける母語話者と学習者作文で使用された接続詞の使用頻度について、One-way ANOVA 分析を行った多重比較の結果、（表 7.10）、学習者（JP2・JP3）と母語話者（CHN）の間に有意差があり（$p < 0.05$）、学習者は母語話者より、各作文ジャンルにおける「接続詞」の使用頻度が著しく高いことを確認した。

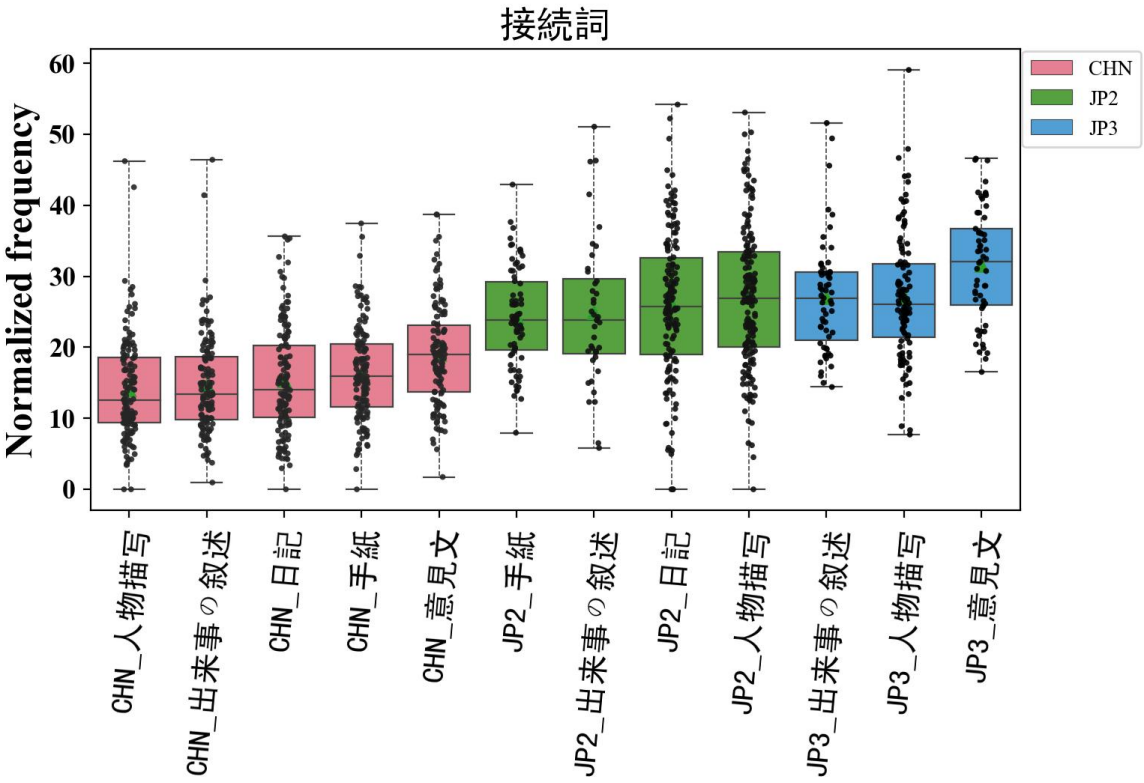

図 7.12　接続詞の Boxplot

表 7.10　母語話者と学習者作文における「接続詞」の多重比較

	接続詞
出来事の叙述	JP3 ≈ JP2 > CHN
人物描写	JP2 ≈ JP3 > CHN
手紙	JP2 > CHN
日記	JP2 > CHN
意見文	JP3 > CHN

注：「>」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

7.3 次元 3 の言語項目

次元 3 を構成する言語項目の相対頻度の平均値を表 7.11 に示した。図 7.13~図 7.14 は、次元 3 における言語項目の相対頻度を用いて描いた Boxplot（箱ひげ図）である。

表 7.11　次元 3 の言語項目の相対頻度（平均値）

Sub corpus	高頻度副詞	時間を表す副詞
CHN_人物描写	58.17	29.04
CHN_出来事の叙述	54.05	27.20
CHN_意見文	45.26	20.58
CHN_手紙	54.22	24.59
CHN_日記	56.17	27.29
JP2_人物描写	47.40	14.23
JP2_出来事の叙述	48.32	15.82
JP2_手紙	52.11	18.46
JP2_日記	50.06	14.00
JP3_人物描写	57.06	17.45
JP3_出来事の叙述	47.75	12.90
JP3_意見文	49.92	13.67

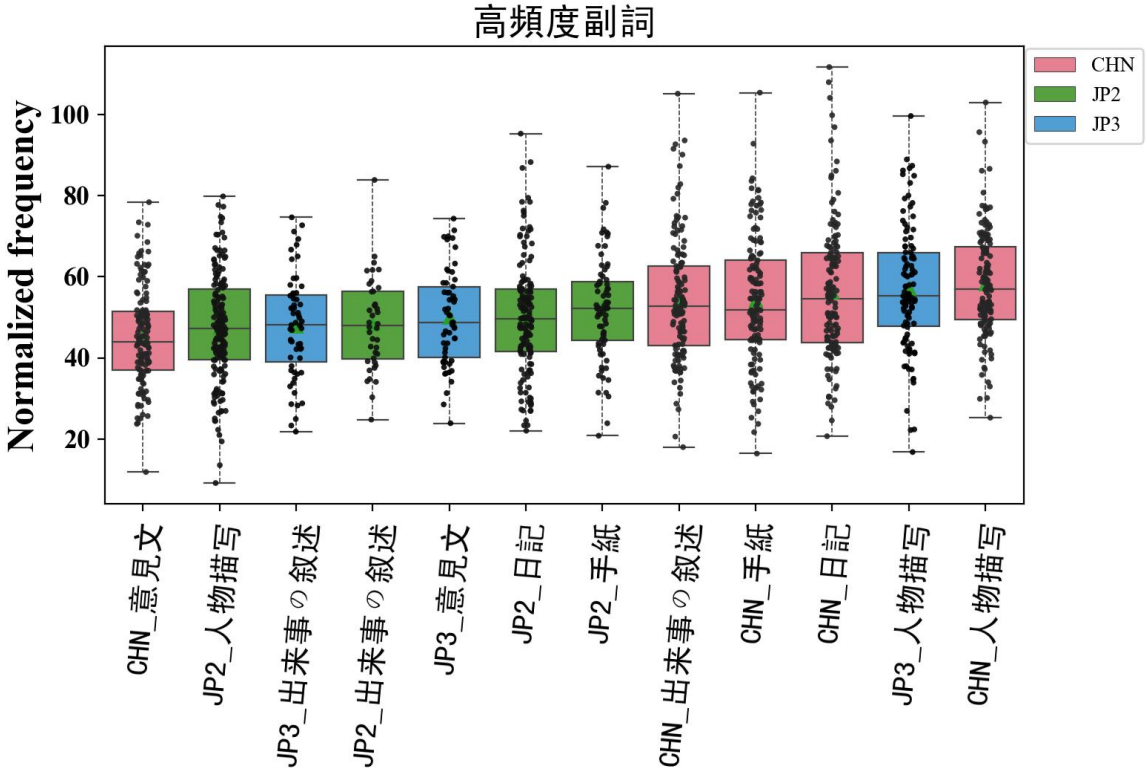

図 7.13　高頻度副詞の Boxplot

図 7.13 に示したように、「CHN_意見文」と「JP3_人物描写」以外に、学習者は母語話者より作文における高頻度副詞の使用頻度が低い傾向が見られる。興味深いのは、「時間を表す副詞」の使用実態である。図 7.14 を見ると、母語話者の方が学習者より全体的に高頻度副詞の使用頻度が高いことが分かった。

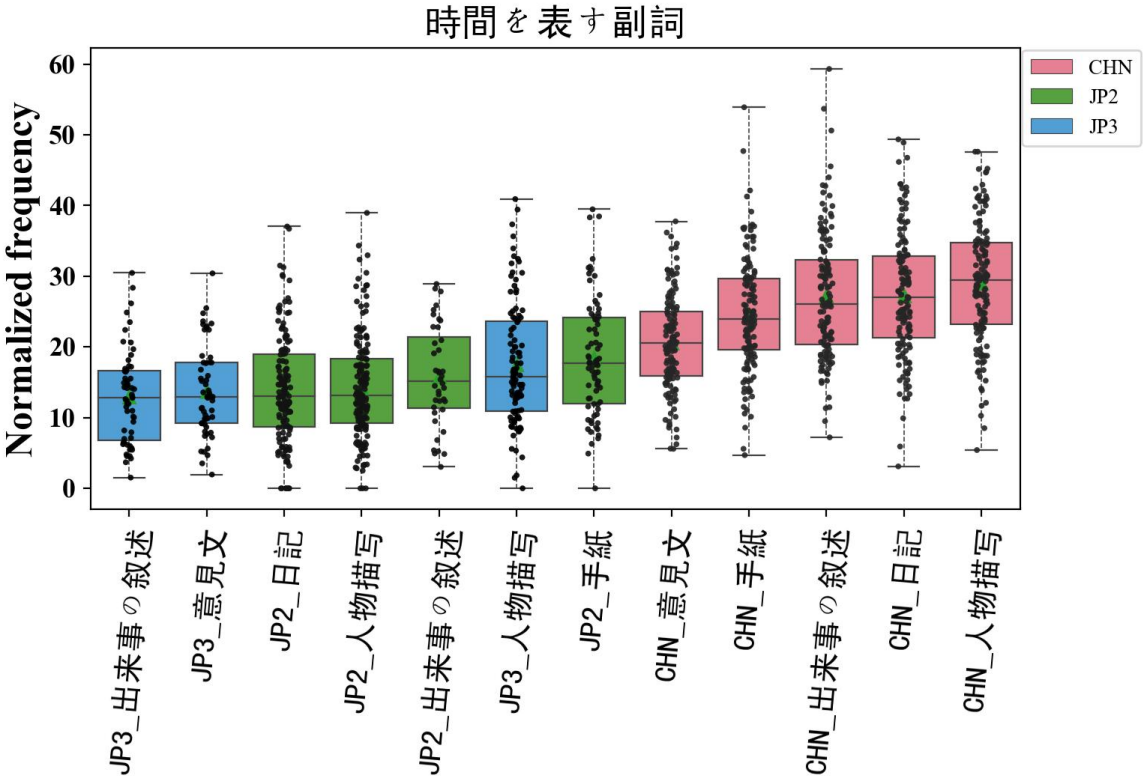

図 7.14　時間を表す副詞の Boxplot

同一ジャンルにおける母語話者作文と学習者作文で使用された「高頻度副詞」と「時間を表す副詞」の頻度について、多重比較を行った。その結果（表 7.12）、「高頻度副詞」より「時間を表す副詞」の方が、学習者作文における出現頻度の傾向はより顕著である。「時間を表す副詞」の使用頻度は、学習者（JP2 ・ JP3）と母語話者（CHN）の間に有意差があり（$p < 0.05$）、母語話者より学習者の方が、各作文ジャンルにおける「時間を表す副詞」の使用頻度が著しく低いことを確認することができた。

表 7.12　母語話者と学習者作文における「高頻度副詞」と「時間を表す副詞」の多重比較

	高頻度副詞	時間を表す副詞
出来事の叙述	CHN ≈ JP3 ≈ JP2	CHN > JP3 ≈ JP2
人物描写	CHN ≈ JP3 > JP2	CHN > JP3 ≈ JP2
手紙	CHN ≈ JP2	CHN > JP2
日記	CHN > JP2	CHN > JP2
意見文	CHN ≈ JP3	CHN > JP3

注：「>」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

7.4 次元 4 の言語項目

次元 4 に属する言語項目は 3 つであり、これら 3 つの言語項目は動詞に関係があるものである。表 7.13 は次元 4 に属する各言語項目の相対頻度の平均値を示している。図 7.15 ~ 図 7.17 は、次元 4 における言語項目の相対頻度を用いて描いた Boxplot（箱ひげ図）である。

表 7.13　次元 4 の言語項目の相対頻度（平均値）

Sub corpus	動作動詞	方向動詞	高頻度動詞
CHN_人物描写	71.59	14.69	115.45
CHN_出来事の叙述	66.62	13.71	109.54
CHN_意見文	48.42	7.51	86.60
CHN_手紙	54.93	9.10	99.53
CHN_日記	59.15	11.81	102.59
JP2_人物描写	50.00	7.18	101.60
JP2_出来事の叙述	56.37	12.82	113.88
JP2_手紙	48.84	11.08	99.99
JP2_日記	60.73	13.56	96.58
JP3_人物描写	46.47	8.69	100.55
JP3_出来事の叙述	55.09	11.79	97.55
JP3_意見文	46.96	5.69	105.53

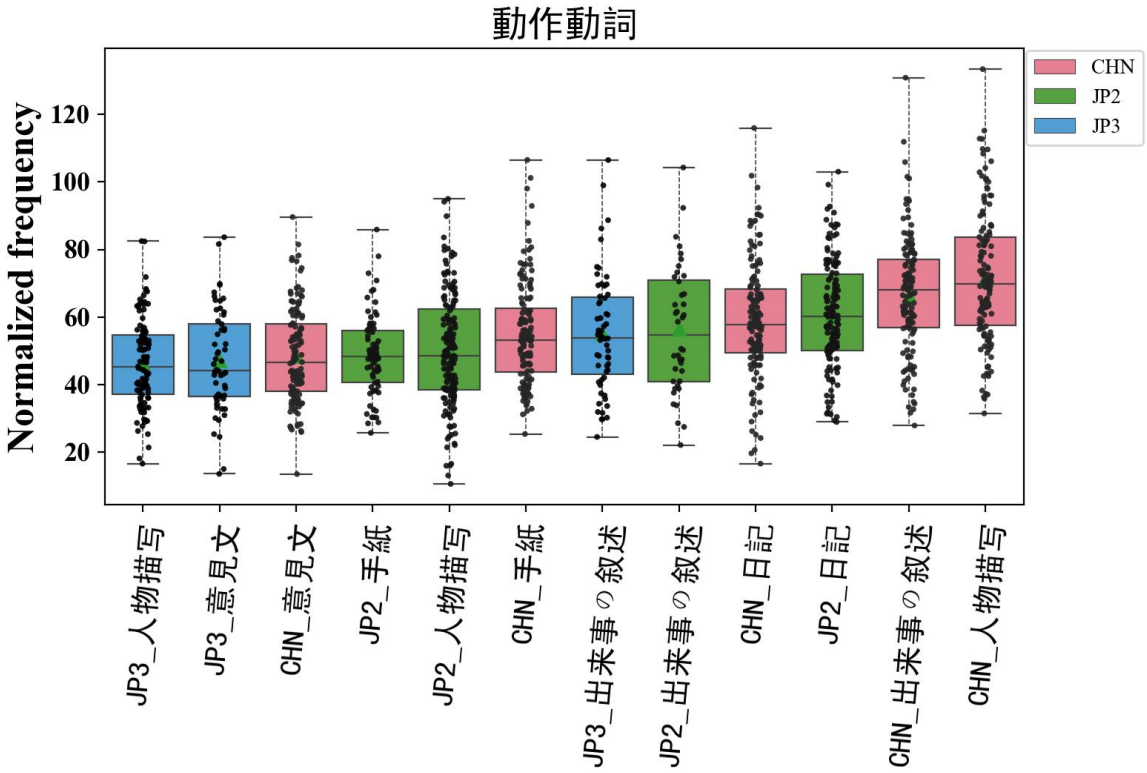

図 7.15　動作動詞の Boxplot

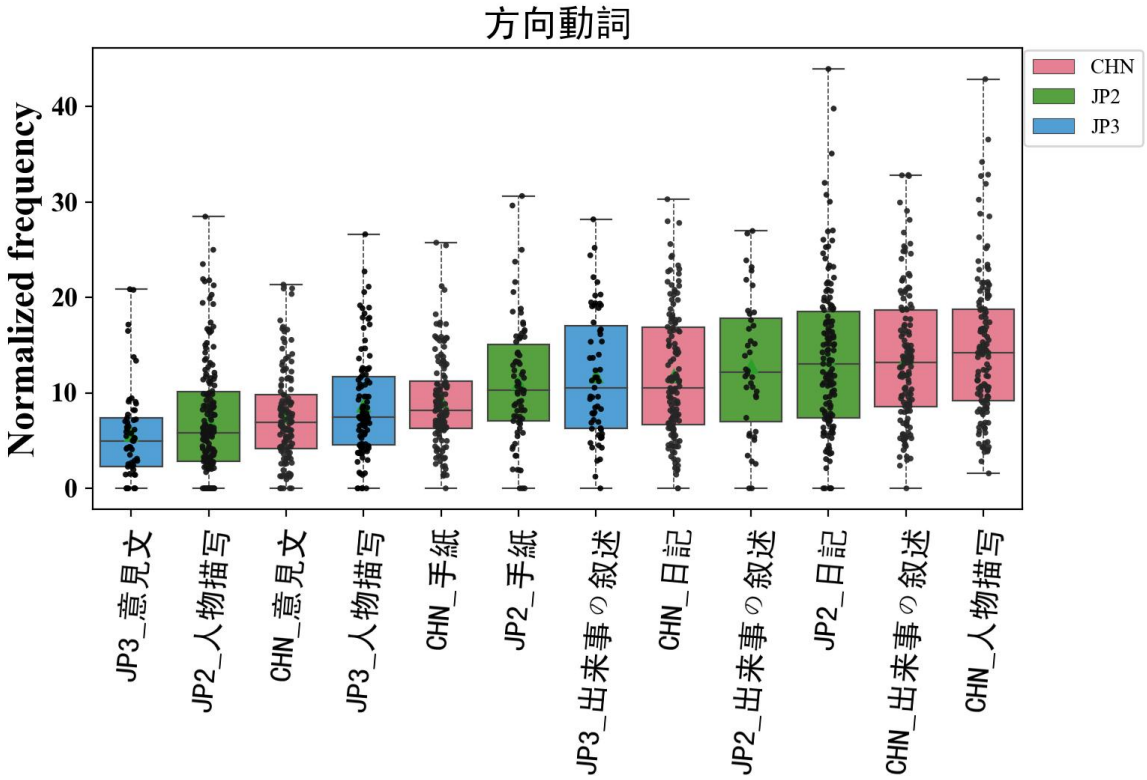

図 7.16　方向動詞の Boxplot

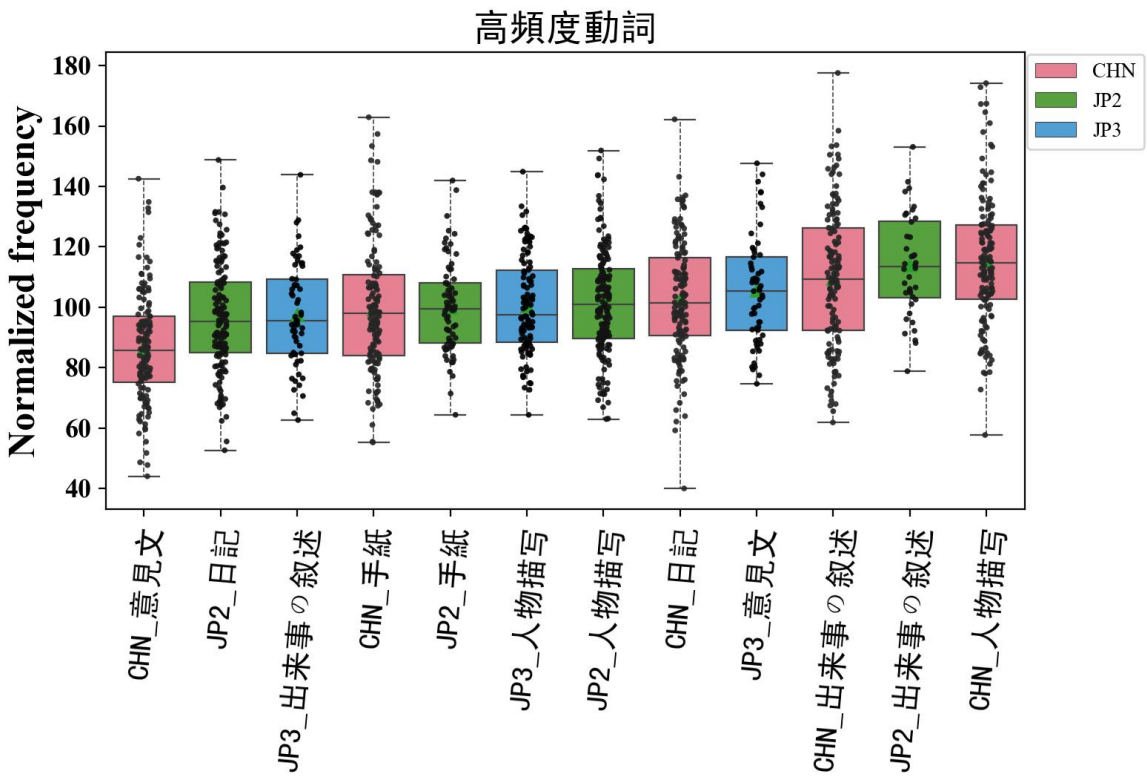

図 7.17　高頻度動詞の Boxplot

これら 3 つの言語項目の相対頻度について、ジャンル間の多重比較を行うことで、以下のような共通の傾向[77]が見られた。

母語話者の作文ジャンルについて、「動作動詞・方向動詞・高頻度動詞」という 3 つの言語項目の使用頻度は、「CHN_人物描写」と「CHN_出来事の叙述」の間（$p > 0.05$）、及び「CHN_日記」と「CHN_手紙」の間（$p > 0.05$）に有意差は認められないものの、「CHN_人物描写/ CHN_出来事の叙述/ CHN_日記/CHN_手紙」と「CHN_意見文」の間に有意差があり（$p < 0.05$）、「CHN_意見文」は他の 4 つのジャンルより、「動作動詞・方向動詞・高頻度動詞」の使用頻度が著しく低いことが確認された。7.15 ~ 図 7.17 からも明らかなように、母語話者作文の 5 ジャンルにおいて、「CHN_意見文」の分布は全体的に低い傾向が見られる。

[77] 多重比較の結果が同一でない部分は記載されていない。例えば、「CHN_出来事の叙述」と「 CHN_日記」の間に、「動作動詞」の使用頻度は有意差があるが、「方向動詞/高頻度動詞」の使用頻度は有意差がない。このような結果の同一でない部分は記載されず、共通の傾向しかまとめていない。

下記の例文（28）と例文（29）は母語話者が書いた人物描写と意見文の抜粋である。例文において、「高頻度動詞」に下線が引かれ（例：“有”、“想”）、「動作動詞」に傍点を付け（例：“打”、“看”）、「方向動詞」はイタリック体（例：“*下来*”、“*去*”）で表示されている。例文（28）では、作文の書き手が、自分の弟の人物像を描写するために、様々な「動作動詞」・「方向動詞」・「高頻度動詞」を用い、弟に関する出来事を叙述することで、弟のやんちゃなキャラクター像を際立たせている。それに対し、例文（29）は、コロナ禍を背景に、オンライン授業のメリットとデメリットについて論じたものである。意見文は事実と自分の意見を繋げて論理的な説明に着眼するのは大切であるため、例文（28）のような人物描写の作文に比べ、例文（29）のような意見文は物語性が少ないので、「動作動詞」や「方向動詞」などの動詞の使用頻度が比較的低いと思われる。

---

(28) 我 家里 有 一个 淘气 无敌 的 弟弟 ， 名叫 小球球 。 小球球 有 一 双 炯炯有神 的 眼睛 ， 眉毛 像 刷子 一样 ， 一 张 能言善辩 ， 爱 拍 “ 马屁 ” 的 小 嘴 巴 ， 一 只 小巧玲珑 的 鼻子 ， 脸蛋 粉红 粉红 的 ， 像 一个 又 大 又 红 的 苹果 。 只要 你 口袋 里面 有 着 钱 ， “ 嗖嗖 ” 地 一下子 ， 他 把 钱 全 都 变成 了 糖果 和 食物 。 记得 有 一 次 ， 妈妈 带 我 和 小球球 *去* 超市 购物 ， 小球球 看见 了 一 把 玩具 小 水枪 ， 边 摇 妈妈 的 手 边 说 ： “ 好 妈妈 ， 你 看 ， 那 把 小 水枪 如果 挂 在 我 身上 肯定 又 帅 又 好看 ， 长大 以后 保护 妈妈 ， 球球 想 要 那个 玩具 ， 你 就 如 了 我 的 心愿 吧 ！ ” 妈妈 听 了 他 的 甜言蜜语 ， 给 小球球 买 了 *下来* 。 *回家* 后 ， 小球球 就 开始 玩弄 *起* 这 把 水枪 。 妈妈 还 想 听 听 那 甜言蜜语 ， 就 走 *去* 说 ： “ 小球球 ， 妈妈 都 给 你 买 了 ， 你 是 不 是 …… ” 妈妈 话 还 没 说 完 ， 只 听见 “ 哇 ” 的 一 声 ， 我 立马 从 房间 跑 出来 ， 我 的 乖乖 ， 小球球 又 拿 小 水枪 打 了 妈妈 一 枪 ， 真 淘气 。

（CHN_人物描写_008, Dimension 4 score: 3.62, 動作動詞: 74.75, 方向動詞: 20.20, 高頻度動詞: 127.27）

(29) 首先 网络 授课 是 把 双刃剑 ， 有 利 也 有 弊 。 各种 授课 app 的 出现 使得 教育 方式 变得 更加 灵活 ， 全面 的 功能 也 让 老师 的 监督 工作 更加 便利 ，

让 同学 们 的 学习 效率 有所 提高 。 直播 授课 的 形式 也 很 好 地 达到 了 课堂 互动 的 目的 。 但是 软件 并非 十全十美 ， 可能 会 因为 网络 延迟 耽误 课堂 进程 ， 也 可能 因为 老师 无法 监视 学生 的 一举一动 使得 部分 学生 课堂 效率 低下 。 同时 ， 网络 交流 无法 像 面对面 交流 那样 方便 且 快速 ， 同学 们 的 问题 不 一定 可以 得到 第一 时间 解决 。 其次 ， 网络 授课 也 是 一 种 考验 。 对于 自觉 的 学生 来说 ， 疫情 期间 是 使 成绩 突飞猛进 的 时间 ， 而 不 自觉 的 学生 则 会 得过且过 ， 仍然 或者 假期 的 懒散 生活 中 ， 虚度 时光 ， 殊不知 努力 的 伙伴 正在 偷偷 赶超 ； 对于 敬业 的 老师 来说 ， 网课 是 一 种 提升 自身 能力 的 方式 ， 而 对于 没有 敬业 精神 的 老师 疫情 期间 他们 将 停滞不前 ； 对于 负责 的 家长 们 来说 ， 自己 的 孩子 会 在 这 段 时间 受益匪浅 ， 自己 也 会 成就感 满满 ， 而 没有 责任感 的 家长 则 会 毫无 所得 。 对于 网络 授课 ， 仁者见仁 ， 智者见智 ， 大家 或许 都 有 自己 的 看法 。

（CHN_意見文_041, Dimension 4 score: -2.56, 動作動詞: 51.46,方向動詞: 2.78, 高頻度動詞: 77.89）

同一ジャンルにおける母語話者作文と学習者作文[78]について、「動作動詞・方向動詞」という 2 つの言語項目の使用頻度は、「CHN_日記」と「JP2_日記」の間（$p > 0.05$）、「CHN_手紙」と「JP2_手紙」の間（$p > 0.05$）、「CHN_意見文」と「JP3_意見文」の間（$p > 0.05$）に有意差がないことを明らかにした（表 7.14）。

興味深いのは、「人物描写」という作文ジャンルである。このジャンルに関しては、「動作動詞・方向動詞・高頻度動詞」の使用頻度はすべて「CHN > JP3 ≈ JP2」の順で、「JP3_人物描写」と「JP2_人物描写」の間（$p > 0.05$）に有意差は認められないものの、学習者（JP3_人物描写 / JP2_人物描写）と母語話者（CHN_人物描写）の間に有意差があり（$p < 0.05$）、学習者（JP3・JP2）が書いた「人物描写」作文における「動作動詞・方向動詞・高頻度動詞」の使用頻度は、母語話者より著しく低いことが確認された。

[78] ここでは、多重比較の結果が同一でない部分も記載されていない。例えば、「出来事の叙述」という作文ジャンルについて、「CHN_出来事の叙述」と「JP3_出来事の叙述 / JP2_出来事の叙述」の間に、「動作動詞」の使用頻度は有意差があるが、「方向動詞」の使用頻度は有意差がない。このような結果の同一でない部分は記載されず、共通の傾向しかまとめて記載していない。

表 7.14　母語話者と学習者作文における「動作動詞・方向動詞・高頻度動詞」の多重比較

	動作動詞	方向動詞	高頻度動詞
出来事の叙述	CHN > JP3 ≈ JP2	CHN ≈ JP3 ≈ JP2	CHN ≈ JP2 > JP3
人物描写	CHN > JP3 ≈ JP2	CHN > JP3 ≈ JP2	CHN > JP3 ≈ JP2
手紙	CHN ≈ JP2	CHN ≈ JP2	CHN ≈ JP2
日記	CHN ≈ JP2	CHN ≈ JP2	CHN ≈ JP2
意見文	CHN ≈ JP3	CHN ≈ JP3	CHN > JP3

注：「>」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

7.5 本章のまとめ

本章では、次元１～4 を構成する有効な各言語項目に対して、同一言語項目における学習者作文と母語話者作文ジャンル間の比較を行った。次章では、第 6 章と第 7 章で得られた結果を踏まえ、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。

第 8 章　中国語教育への応用

本章では、まず、第 6 章と第 7 章の考察によって得られた結果を踏まえ、学習者の作文における特徴と問題点をまとめる。そして、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に関する特徴及び問題点の要因を解明する。最後に、中国語教育、特に、中国語作文指導の改善へ向けての提言を行う。

8.1 学習者の作文における特徴と問題点

第 6 章と第 7 章はそれぞれ 、次元と言語項目の側面から、同一ジャンルにおける母語話者作文と学習者作文の比較を行った。下記に示した表 8.1~8.4 は、各次元の次元スコアと各次元を構成する言語項目の相対頻度について、6.2 節と 7.1～7.4 節から得られた同一ジャンルにおける母語話者作文と学習者作文間の多重比較の結果をまとめたものである。表 8.1~8.4 において、各表の最後の列は、同一作文ジャンルの母語話者作文と学習者作文を各次元で比較した結果（6.2 節と表 6.19 を参照されたい）であり、他の列は、各次元を構成する言語項目の出現頻度について、同一ジャンにおける母語話者作文と学習者作文間の比較結果（7.1～7.4 節を参照されたい）である。

(1) 次元 1 の機能ラベルから見られるように（表 8.1）、学習者の作文における産出語彙の豊富さは母語話者より著しく低く、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いる。さらに、次元 1 を構成する言語項目の使用頻度について、学習者作文と母語話者作文の違いを比較した結果（表 8.1）、以下の 2 点が明らかになった。

① 学習者（JP2・JP3）の作文における「高難度語彙」と「非 HSK 語彙」の使用頻度、及び「平均文長」と「語彙多様性」は、母語話者（CHN）の作文より著しく低い。

② 「出来事の叙述」を除くその他の 4 つのジャンルにおいて、「低難度語彙」と「一人称単数代名詞“我”」使用頻度は、母語話者（CHN）作文より学習者（JP2・JP3）作文の方が著しく高い。なお、「出来事の叙述」における「低難度語彙」と「一人称単数

代名詞“我”」使用頻度は、中級学習者（JP2）が母語話者（CHN）より著しく高いものの、上級学習者（JP3）と母語話者（CHN）の間には、有意差が認められなかった。

従って、JP3（上級学習者）の「出来事の叙述」という作文ジャンルでは、「低難度語彙」と「一人称単数代名詞“我”」の使用に対するコントロールは、ほぼ母語話者（CHN）のレベルに達しているが、それとは別に、学習者（JP2 ・JP3）の作文ジャンルにおいて、「低難度語彙」と「一人称単数代名詞“我”」が著しく過剰に使用されたことを明らかにした。一方、学習者の 5 ジャンルの作文には、「高難度語彙」と「非 HSK 語彙」の過少使用、センテンスの長さが短い、語彙の多様性が低いという問題点も見られた。

表 8.1　次元 1：産出語彙の豊富さ vs. 産出語彙の限定性

ジャンル	「産出語彙の豊富さ」を表す言語項目				
	高難度語彙	語彙多様性	非 HSK 語彙	平均文長	韻律双音節語
出来事の叙述	CHN > JP3 > JP2	CHN > JP2 > JP3	CHN > JP3 ≈ JP2	CHN > JP3 ≈ JP2	CHN ≈ JP3 ≈ JP2
人物描写	CHN > JP3 ≈ JP2	CHN > JP3 > JP2	CHN > JP3 ≈ JP2	CHN > JP3 > JP2	CHN ≈ JP3 ≈ JP2
手紙	CHN > JP2	CHN > JP2	CHN > JP2	CHN > JP2	CHN ≈ JP2
日記	CHN > JP2	CHN > JP2	CHN > JP2	CHN > JP2	CHN ≈ JP2
意見文	CHN > JP3	CHN > JP3	CHN > JP3	CHN > JP3	CHN ≈ JP3

ジャンル	「産出語彙の限定性」を表す言語項目		次元 1：産出語彙の豊富さ vs. 産出語彙の限定性
	一人称単数代名詞“我”	低難度語彙	
出来事の叙述	JP2 > JP3 ≈ CHN	JP2 > JP3 ≈ CHN	CHN > JP3 > JP2
人物描写	JP2 > JP3 > CHN	JP3 ≈ JP2 > CHN	CHN > JP3 > JP2
手紙	JP2 > CHN	JP2 > CHN	CHN > JP2
日記	JP2 > CHN	JP2 > CHN	CHN > JP2
意見文	JP3 > CHN	JP3 > CHN	CHN > JP3

注：「>」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

(2) 次元 2 の機能ラベルから分かるように（表 8.2）、学習者の作文における状態状況表現の適切さは母語話者より著しく低い。次元 2 を構成する言語項目の使用頻度について、学習者作文と母語話者作文の違いを比較した結果、以下の 2 つの傾向が見られた。

① 学習者（JP2・JP3）の作文における「時態助詞“着”」の使用頻度は、母語話者（CHN）の作文より著しく低い。

② 「ぼかし言葉・程度を表す副詞・接続詞」使用頻度及び「単語の平均長さ」について、母語話者（CHN）作文より学習者（JP2・JP3）作文の方が著しく高い。

しかし、「単語の平均長さ」については、7.2 節ですでに解明したように、学習者の作文における日本語の人名・団体名・地名の単語の長さは、中国語母語話者の作文における単語よりも一般的に長いからであるため、母語話者より学習者の方が、作文における単語の長さが全体的に長い傾向が見られる。この傾向は、日本人中国語学習者の作文としての特徴の一つであり、作文上の問題とは言えない。従って、本節最後のまとめの部分では、「単語の平均長さ」という言語項目については、議論しないことにする。

表 8.2 次元 2：状態状況表現の適切さ

ジャンル	言語項目				
	時態助詞 “着”	ぼかし 言葉	程度を 表す副詞	接続詞	単語の 平均長さ
出来事の叙述	CHN > JP3 ≈ JP2	JP3 ≈ JP2 > CHN	JP3 ≈ JP2 > CHN	JP3 ≈ JP2 > CHN	JP3 > JP2 > CHN
人物描写	CHN > JP3 ≈ JP2	JP2 ≈ JP3 > CHN	JP2 ≈ JP3 > CHN	JP2 ≈ JP3 > CHN	JP2 > JP3 > CHN
手紙	CHN > JP2	JP2 > CHN	JP2 > CHN	JP2 > CHN	JP2 > CHN
日記	CHN > JP2	JP2 > CHN	JP2 > CHN	JP2 > CHN	JP2 > CHN
意見文	CHN > JP3	JP3 > CHN	JP3 > CHN	JP3 > CHN	JP3 > CHN

ジャンル	次元 2： 状態状況表現の適切さ
出来事の叙述	CHN > JP2 > JP3
人物描写	CHN > JP3 ≈ JP2
手紙	CHN > JP2
日記	CHN > JP2
意見文	CHN > JP3

注：「>」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

(3) 次元 3 の機能ラベルを見ると（表 8.3）、「意見文」以外に、学習者の作文における副詞の修飾は母語話者より著しく低い。次元 3 を構成する言語項目の使用頻度について、学習者作文と母語話者作文の違いを比較した結果、学習者（JP2・JP3）の作文において、特に「時間を表す副詞」の使用頻度は、母語話者（CHN）の作文より著しく低いことを確認した。

表 8.3　次元 3：副詞の修飾

ジャンル	言語項目		次元 3：副詞の修飾
	高頻度副詞	時間を表す副詞	
出来事の叙述	CHN ≈ JP3 ≈ JP2	CHN > JP3 ≈ JP2	CHN > JP3 ≈ JP2
人物描写	CHN ≈ JP3 > JP2	CHN > JP3 ≈ JP2	CHN > JP3 > JP2
手紙	CHN ≈ JP2	CHN > JP2	CHN > JP2
日記	CHN > JP2	CHN > JP2	CHN > JP2
意見文	CHN ≈ JP3	CHN > JP3	CHN ≈ JP3

注：「>」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

(4) 次元 4 の機能ラベルと次元 4 を構成する言語項目から見れば（表 8.4）、各ジャンルにおける学習者作文と母語話者作文間の比較について、最も顕著なのは、「人物描写」という作文ジャンルである。なぜなら、「人物描写」作文における「動作動詞」・「方向動詞」・「高頻度動詞」の使用頻度は、学習者間（JP2 と JP3）には有意差が認められないものの、母語話者（CHN）より学習者（JP2 或は JP3）の方が著しく低い（$p < 0.05$）という共通の傾向が見られるからである。従って、学習者（JP2・JP3）の「人物描写」作文では、「動作動詞」・「方向動詞」・「高頻度動詞」の使用が不十分である問題点がある。

表 8.4　次元 4：動作の描写

ジャンル	言語項目			次元 4：動作の描写
	動作動詞	方向動詞	高頻度動詞	
出来事の叙述	CHN > JP3 ≈ JP2	CHN ≈ JP3 ≈ JP2	CHN ≈ JP2 > JP3	CHN ≈ JP2 > JP3
人物描写	CHN > JP3 ≈ JP2	CHN > JP3 ≈ JP2	CHN > JP3 ≈ JP2	CHN > JP3 ≈ JP2
手紙	CHN ≈ JP2	CHN ≈ JP2	CHN ≈ JP2	CHN ≈ JP2
日記	CHN ≈ JP2	CHN ≈ JP2	CHN ≈ JP2	CHN ≈ JP2
意見文	CHN ≈ JP3	CHN ≈ JP3	CHN > JP3	CHN ≈ JP3

注：「>」マークは有意差あり（$p < 0.05$）、「≈」マークは有意差なし（$p \geqq 0.05$）を意味している。

学習者の作文における特徴と問題点をまとめると、学習者の中国語作文には、大きく分けて下記のような 3 つの特徴と問題がある。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著しく低い。三、「人物描写」という作文ジャンルでは、学習者の方が、「動作動詞」・「方向動詞」・「高頻度動詞」の使用が明らかに不十分である問題点があることも解明した。

8.2 在日中国語教育向けの提言

8.2.1 言語項目の過剰使用について

8.1 節から明らかなように、学習者が母語話者より全体的に「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」を過剰に使用している傾向が見られる。ある言語項目の過剰使用は、必ずしも誤用を意味するものではないが、適切に使わないと、文章のまとまりや読みやすさが損なわれてしまう場合がある。例えば、馬（2017）が述べているように、一人称代名詞の過剰使用は、談話の一貫性に影響を与えるだけでなく、書き言葉としての書面性を損なう可能性がある。

(1) 「一人称単数代名詞“我”」の過剰使用について

本研究では、学習者作文における「一人称単数代名詞“我”」が過剰に使用している結果は、先行研究（高 1996; 馬 2017; 徐 2021）の結果と一致している。

L2 中国語ライティングの研究分野において、高（1996）は、留学生の文章表現に“我”（私）という代名詞が過剰に使われている問題があると論じている。馬（2017）では、中国語母語話者の書き言葉より、中国語学習者の HSK 作文[79]の方が一人称代名詞“我”の使用頻度が高いことを明らかにした。 徐（2021）は、タイ人中国語上級学習者が書いた叙述文における一人称代名詞の出現頻度が、中国語母語話者より著しく高いことを証明した。一人称代名詞の過剰使用は、談話の一貫性に影響を与えるだけでなく、書き言葉としての書面性が損なわれてしまうと述べられている（馬 2017）。

また、L2 英語ライティングの文章表現において、一人称代名詞を多用する傾向があることを示す研究が多く行われてきた（Petch-Tyson 2014; 盧 2007[80]; 趙・尚 2012[81]等）。例えば、日本人学生の英作文において一人称代名詞を多用する研究（石川 2008; McCrostie 2008）が見られる。石川(2008: 220)では、日本人学生が英作文において人称代名詞 *I* や *I think* という表現を多用することを報告している。McCrostie(2008)は、日本人学習者が英作文において個人的な見解や意見を表現するために *I think* をよく使用することを示している。

本研究では、作文の文脈における一人称単数代名詞“我”の右側にある最初の単語を集計した。上位 10 語の共起パターン（例えば、“我 觉得”/“我 很”）は、図 8.1 に示した Network（ネットワーク）可視化図で表示されている。

---

[79] HSK 作文は、1992 年から 2005 年にかけて HSK5 級試験を受験した受験者によって書かれた作文を指す。

[80] 卢海燕(2007). 大学英语写作中人称指代使用的量化对比研究. 河南师范大学学报: 哲学社会科学版, 34(4), 157-158.

[81] 赵小东・尚晓华(2012). 从人称代词角度考察大学生作文的口语化表现. 外语学刊, (6), 119-123.

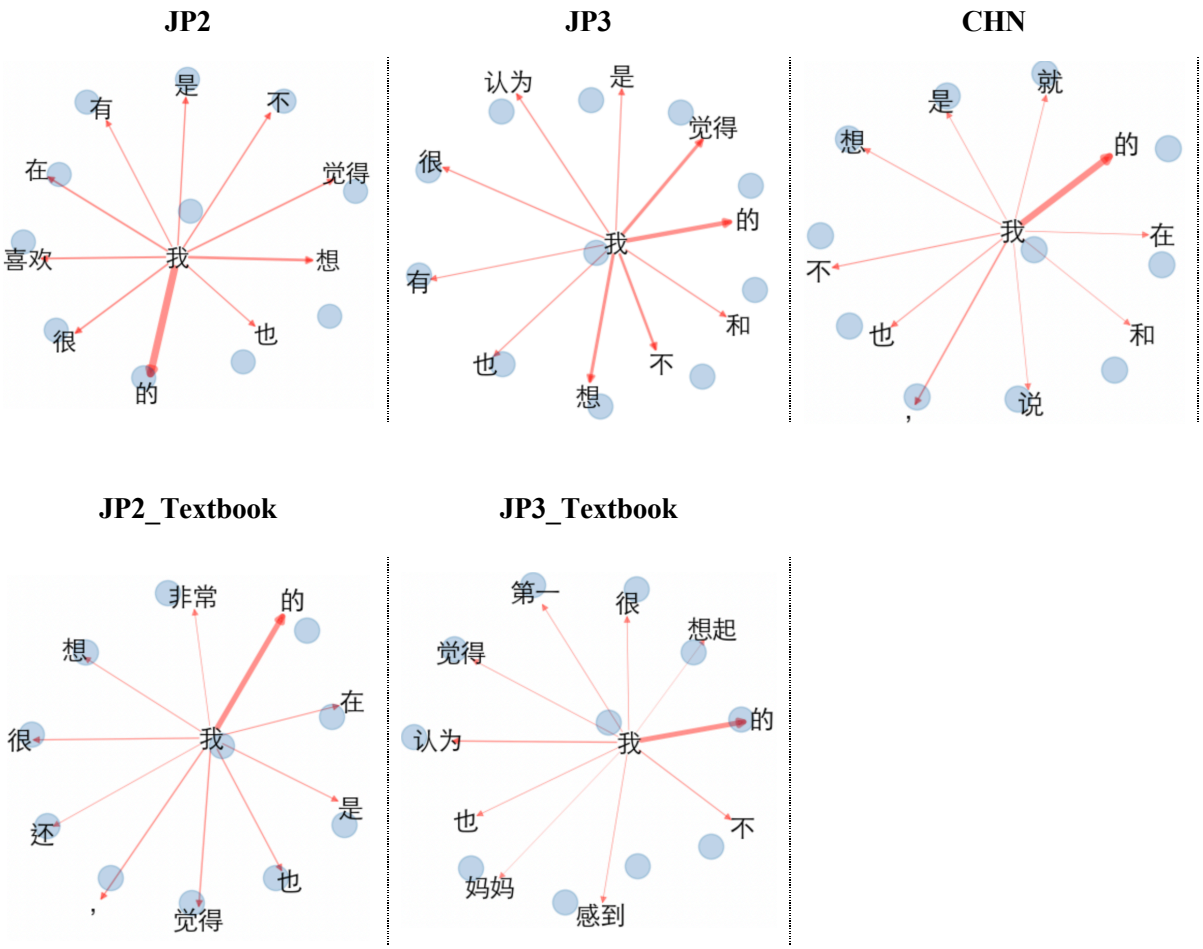

図 8.1　“我 X”：共起パターンの Network 可視化図（上位 10 語）

図 8.1 に示した共起パターンからわかるように、“我的”（和訳：私の）、“我也”（和訳：私も）、“我是”（英訳：I am~）、“我不”（英訳：I don't~）“我想”（英訳：I want~）という 5 つの共起パターンは、学習者（JP2、JP3）と母語話者（CHN）の Network 図に現れていることが分かった。

しかし、図 8.1 に示したように、“我觉得”（英訳：I think~）と“我很”（英訳：I'm very~）という 2 つのパターンが際立っていると言える。例文（30）～（33）は、学習者の作文文脈における“我很” と“我觉得”の抜粋を示している。この 2 つのパターンは、学習者（JP2、JP3）や教科書（JP2_Textbook、JP3_Textbook）の Network 図には現れているが、母語話者（CHN）の上位 10 個の共起パターンには出現していない。また、上位 10 語の共起パターンにおいて、 “我认为”（英訳：I believe that~ / I think ~）は JP3 と JP3_Textbook に出現しているが、JP2 と JP2_Textbook 及び CHN に出現していない。従って、学習者の作文における一人称代名詞とよく共起した表現の一部は、主に教科書

の影響を受けていると言える。なぜなら、教科書によく現れる固定表現は、学習者が目にする機会が多いため、学習者の産出したパターンにも多く含まれていると考えられるからである。

例文：

(30) 吃 午饭 的 以后 ， 我 上 下午 的 课 。 我 学习 英国 政治 。 这 课 难 一点儿 ， 可是 很 有意思 。 我 想 学习 世界 各国 。 上课 的 以后 ， 我 参加 合唱团 。 多 新生 也 参加 活动 ， 我 很 高兴 。 参加 社团 的 以后 ， 我 预习 和 复习 汉语 。 最后 我 睡觉 睡 得 早 。 今天 我 很 累 可是 过 很 有 意义 的 生活 。

(JP2_出来事の叙述_006, Dimension 1 score: - 9.35, 一人称単数代名詞“我”: 88.77)

(31) 箕面校区 的 女 学生 比 丰中 更 多 。 可是 听说 吹田校区 的 女 学生 比 丰中 很少 。 我 觉得 我 在 上大学 ， 只是 在 不同 的 校园 。 回顾 第一 天 ， 校园 比 去年 变 了 ， 有 很多 不耐烦 的 情绪 。 我 觉得 在 去 一个 新 环境 时 ， 我 需要 比 平时 更 仔细 地 准备 。

(JP2_出来事の叙述_018, Dimension 1 score: - 4.91, 一人称単数代名詞“我”: 65.49)

(32) 从 我 上 高中 时 起 ， 我们 不 住 在 一起 ， 所以 我 离 她 住 已经 5 年 多 了 。 我 和 她 住 的 时候 ， 每天 跟 她 吵架 了 。 可是 ， 最近 的 话 ， 没 有 吵 过 。 小时候 ， 我 觉得 她 非常 可憎 ， 现在 觉得 可爱 一点儿 。

(JP3_人物描写_007, Dimension 1 score: - 3.71, 一人称単数代名詞“我”: 38.78)

(33) 由于 学系 和 社团 活动 都 是 一样 的 ， 我们 每天 见面 ， 有 很多 共同 话题 。 我 很 高兴 ， 因为 没 想到 在 大学 里 能 交 到 这样 的 朋友 。 最 难忘 的 回忆 是 他 访问 我 家 的 时候 。 这 是 因为 我 觉得 我 和 她 的 友谊 加深 了 。

(JP3_人物描写_056, Dimension 1 score: - 2.79, 一人称単数代名詞“我”: 51.09)

興味深いのは、学習者がよく使用する共起パターン「“我觉得”（英訳：I think~）」である。英語では、「*I think*」は話し手の立場を最も直接的に示す言葉の一つである（Biber *et al.* 1999: 976）。日本人学生は、英作文で個人的な意見を表現するために「*I*

think」を多用するため、優柔不断な議論になってしまった（Oi 1999）と述べられている。Natsukari（2012）は、「*I think*」の多用は、日本語の「思う」の直訳と、日本語文脈に関わる優柔不断な書き方と関係があると論じている。日本語の「思う」は主に意見や不確かさを表現するために使われ、「*I think*」と「思う」の主な定義が似ているため、この類似性から日本人英語学習者は英文の中で「*I think*」を使って個人の意見を表現しやすくなってしまうと言える。また、石川（2008: 221）では、日本大学生の日本語作文における全構成語彙の中で、「思う」が 2 番目に多い語彙であることを述べており、日本の学生が「思う」を多用するのは、日本語母語の負の転移と関係があるのではないかと指摘している。従って、本研究では、日本人中国語学習者の作文において「“我觉得”（英訳：I think~）」という表現を多用している原因は、教科書の影響に加え、「思う」という日本語の直訳と日本語母語の負の転移に帰されるのではないかと考えられる。

学習者作文において、「一人称単数代名詞“我”」を過剰に使用した原因は、上記に述べた 2 つの原因（教科書の影響と日本語母語の負の転移）に加え、もう一つの原因は、学習者が人称代名詞を省略する方策がまだ習得していないことが強く疑われる。

次の例文(34)は、学習者が書いた叙述文の抜粋である。部活の人が新型コロナウイルスに感染したため、学園祭で部活全員がパフォーマンスを披露できない可能性があると言われたことについて、学習者は作文で自分の失望感を表現している。

---

(34) 我 没有 去 过 学园祭 ， 所以 期待 了 很 久 了 。我 真 的 很 悲伤 。 我 得 了 灰色 的 人生观 。 我 的 心情 从 高兴 到 难过 变 了 。 那 时候 ， 我 忘 了 好 几个 幸福 ， 集中 不幸 。 我 过 了 不安 非 观 的 几 天 。 几 天 后 我 觉得 我 要 数 幸福 ， 不 要 数 不幸 。（JP2_出来事の叙述_008）

直訳：私は学園祭に行ったことがないので、ずっと楽しみにしていました。私は本当に悲しかったです。私は灰色の人生観を手に入れました。私の気分は楽しいから悲しいことに変わりました。その時、私はいくつかの幸福を忘れ、不幸を集中しました。私は不安的悲観的な日々を過ごしました。数日後、私が、不幸ではなく、幸福を数えなければならないと私は感じました。

(35) 我 先 是 一 惊 ， Ø之后 开心 无比 ， Ø同时 又 信心 倍增 。 Ø用 同样 的 方法 ， Ø在 爸爸 的 陪伴 下 ， Ø绕 着 小区 骑 了 好几 圈 。（CHN_出来事の叙述_078）

直訳：私はまず驚き、次に嬉しさと自信が同時に湧いてきました。同じような方法で、父に連れられて何度も乗り回した。

上記の例文(34)に示したように、学習者は、文脈に応じて**ゼロ代名詞**[82]（中国語：零形代词；英語：zero pronoun）を使ったり人称代名詞を省略したりすることが苦手なため、文章表現に短いセンテンスを積み上げ、センテンス毎に“**我**”（私）を言及する傾向がある。それに対して、中国語母語話者は、初めて自転車に乗れるようになったときの気持ちを表現する際(例文 35)、トピックチェーン(中国語:话题链；英語:Topic chain)のある節に一人称代名詞“**我**”（私）を使ったら、同じトピックチェーンの他の節には自然に“**我**”（私）を省略してゼロ形（記号：Ø）の主語を使っている。

王・王（2019）[83]の指摘によると、人称代名詞を多用することにより、叙述文は常に人称代名詞を主語とし、単文構造になってしまうという問題が生じている[84]。上記の例文(34)において必要ではない人称代名詞を省略し、短いセンテンスを長い文に統合して誤字・脱字を修正した後の自然な文は以下のようになる。

---

[82] ゼロ代名詞の解釈について、日本中国語学会による『中国語学辞典』の P362 における内容を引用すると次のようになる。
ゼロ代名詞（零形代词 zero pronoun）：文中のありそうな箇所に代名詞が現れない時、形を持たずに存在すると考える理論上の代名詞的要素。主には人称代名詞、時に指示詞がゼロ代名詞に置き換わる。空集合を表す「Ø」や「e」の記号で示されることが多い。省略という捉え方もあるが、義務的でない要素が現れないのは省略とはしないとする立場もあり、そのような立場に立ってもゼロ代名詞という捉え方は有効である。ゼロ代名詞は、主語、目的語、連体修飾語の位置に現れるが、主語の位置にもっとも現れやすい（→代名詞主語省略）。

[83] 王保健・王宝平(2019). 英语写作人称代词过度使用及其教学对策. 内蒙古师范大学学报(教育科学版), 32(2), 99-103.

[84] 王・王(2019)：“人称代词的过度使用……造成记叙文写作总是以人称作句子的主语，句子结构单一”。

(34)_修正：我 没有 去 过 学园祭 ， 所以 期待 了 很 久 了 ，但 现在 真 的 很 悲 伤 。 那 时候 ， 我 忘 了 自己 所 拥有 的幸福 ， 只 看 到 了 自己 的 不幸 。 在 度 过 了 不安 悲观 的 几 天 后 ，我 觉得 应该 要 多 看 幸福 的 一面， 不 要 只 关注 于 不幸 。（JP2_出来事の叙述_008）

訳：私は学園祭に行ったことがないので、ずっと楽しみにしていたのですが、今は本当に悲しいです。その時、私はこれまでの幸せを忘れ、自分の不幸ばかりを見ていました。不安で悲観的な数日間を過ごしていた後、不幸ばかりに目を向けるのではなく、もっと幸せなことに目を向けるべきだと私は思いました。

学習者の作文における一人称代名詞を多用する傾向に対して、徐（2021）の提言により引用すると下記のようになる。

徐（2021）：

“在对外汉语写作教学中，应重视第一人称代词表达策略的训练……应意识到学习者在汉语写作中存在滥用第一人称代词的问题，可通过进行第一人称代词的省略、纠错练习，有意识地纠正学习者言必称‘我’的习惯……”

つまり、学習者向けの中国語作文教育において、一人称代名詞の表現ストラテジーの練習を重視すべきである。中国語作文指導を行う際、学習者の中国語作文では一人称代名詞を過剰に使用する傾向があるという問題を認識する必要があり、一人称代名詞の省略や訂正を学習者に練習させることを加え、学習者のいつも“**我**”（私）を言及する習慣を意識的に矯正することが必要である。

(2) 「低難度語彙」の過剰使用について

図 8.2 は、第 4 章で紹介したコーパスの Wordlist 検索機能を用いて作成した低難度語彙の Word cloud 可視化図である。Word cloud で可視化すると、学習者（JP2、JP3）と母語話者（CHN）サブコーパスにおける低難度語彙の全体像を把握することができる。図 8.2 を見ると、学習者（JP2、JP3）の Word cloud は比較的に近似しているが、学習者と母語話者の間には、全体的に大きな違いが見られる。

JP2

JP3

CHN

図 8.2　低難度語彙の Word cloud

表 8.5　語彙別の低難度語彙：頻度と相対頻度

JP2 (RTTR: 1.09)				JP3 (RTTR: 1.30)				CHN (RTTR: 0.74)			
	語彙	頻度	相対頻度		語彙	頻度	相対頻度		語彙	頻度	相対頻度
	Total	70045	363.18		**Total**	47933	329.98		**Total**	160798	308.34
1	我	9134	47.36	1	的	6732	46.34	1	的	24200	46.41
2	的	8024	41.60	2	我	4363	30.04	2	我	11945	22.91
3	了	3003	15.57	3	是	1809	12.45	3	了	9237	17.71
4	是	2377	12.32	4	了	1590	10.95	4	是	5840	11.20
5	在	1825	9.46	5	在	1463	10.07	5	一	5081	9.74
6	很	1766	9.16	6	不	1267	8.72	6	在	4798	9.20
7	和	1309	6.79	7	很	1113	7.66	7	不	4303	8.25
8	有	1164	6.04	8	我们	1063	7.32	8	你	3273	6.28
9	去	1164	6.04	9	人	1022	7.04	9	着	3049	5.85
10	不	1120	5.81	10	有	901	6.20	10	我们	2784	5.34
11	人	896	4.65	11	她	870	5.99	11	有	2562	4.91
12	喜欢	890	4.61	12	和	845	5.82	12	就	2473	4.74
13	所以	838	4.35	13	去	724	4.98	13	也	2151	4.12
14	一	834	4.32	14	他	701	4.83	14	他	2141	4.11
15	也	804	4.17	15	也	697	4.80	15	人	1898	3.64
16	我们	801	4.15	16	所以	633	4.36	16	这	1896	3.64
17	想	774	4.01	17	一	587	4.04	17	上	1888	3.62
18	好	739	3.83	18	想	499	3.44	18	您	1879	3.60
19	你	735	3.81	19	时候	491	3.38	19	都	1813	3.48
20	看	734	3.81	20	但是	452	3.11	20	说	1736	3.33
	…	…	…		…	…	…		…	…	…
288	右边	1	0.01	284	读	1	0.01	296	出租车	1	0.00

表 8.5 は語彙別の低難度語彙（上位 20 語）の頻度と相対頻度を示している。表 8.5 からわかるように、上位 20 語の低難度語彙において、JP2 と JP3 には、「一人称単数代名詞“我”（訳：私）」、“很”（訳：とても）のような程度を表す副詞、“和”（訳：と）・“所以”（訳：だから）のような接続詞が高頻度で表れている一方、CHN には、時態助詞“着”（訳：～している）が高頻度で出現している。「低難度語彙」は次元 1 に属する言語項目の 1 種であるが、これらの低難度語彙に対応する「程度を表す副詞」、「接続詞」、「時態助詞“着”」も次元 2 を構成する言語項目の語彙である。

(3) 「ぼかし言葉」・「程度を表す副詞」の過剰使用について

「程度を表す副詞」に属する語彙は「ぼかし言葉」の一部である。第 4 章で紹介した作文コーパスの検索機能を用い、JP2・JP3・CHN 作文コーパス、及び JP2 教科書（JP2_Textbook）、JP3 教科書（JP3_Textbook）の例文における「ぼかし言葉」と「程度を表す副詞」の出現頻度を集計した。

表 8.6 は各サブコーパスにおける「ぼかし言葉」の粗頻度と相対頻度を高い順に示した結果である。「ぼかし言葉」の異なり語数が多いため、表 8.6 主に上位 10 語のぼかし言葉の表現を示している。

表 8.6　語彙別の「ぼかし言葉」の頻度集計（上位 10 語）

JP2				JP3				CHN			
No.	Words	Freq	R_Freq	No.	Words	Freq	R_Freq	No.	Words	Freq	R_Freq
	Tokens	6148	32.78		Tokens	4778	33.83		Tokens	10344	19.38
1	很	1725	9.20	1	很	1097	7.77	1	多	1127	2.11
2	常	553	2.95	2	很多	377	2.67	2	很	1030	1.93
3	很多	506	2.70	3	常	329	2.33	3	几	534	1.00
4	多	445	2.37	4	可以	243	1.72	4	些	532	1.00
5	我觉得	267	1.42	5	多	229	1.62	5	将	494	0.93
6	可以	242	1.29	6	我觉得	227	1.61	6	可以	486	0.91
7	太	172	0.92	7	应该	138	0.98	7	太	454	0.85
8	常常	137	0.73	8	几	123	0.87	8	少	390	0.73
9	将	129	0.69	9	太	118	0.84	9	常	383	0.72
10	有时	119	0.63	10	可能	114	0.81	10	总是	325	0.61
…	…	…	…	…	…	…	…	…	…	…	…
86	越发	1	0.01	81	个别	1	0.01	110	绝大	1	0.00

JP2_Textbook				JP3_Textbook			
No.	Words	Freq	R_Freq	No.	Words	Freq	R_Freq
0	Tokens	299	29.35	0	Tokens	211	31.53
1	很	55	5.40	1	很	36	5.38
2	很多	21	2.06	2	多	18	2.69
3	常	21	2.06	3	很多	17	2.54
4	多	18	1.77	4	可以	16	2.39
5	一些	11	1.08	5	少	13	1.94
6	可以	11	1.08	6	太	9	1.34
7	几	9	0.88	7	些	7	1.05
8	差不多	8	0.79	8	不少	7	1.05
9	许多	8	0.79	9	刚	5	0.75
10	我觉得	8	0.79	10	有些	5	0.75
…	…	…	…	…	…	…	…
62	近来	1	0.10	55	极大	1	0.15

注：Tokens は総語数、「Freq」は粗頻度、「R_Freq」は 1,000 字あたりの相対頻度である。

図 8.3 は、JP2・JP3・CHN コーパス、JP2 教科書(JP2_Textbook)、JP3 教科書(JP3_Textbook)の例文における「程度を表す副詞」の全体像を Word Cloud で可視化している。Word Cloud では、テキストに高い頻度で表れている語彙が大きいフォントで表示される（小林 2017: 119）。

図 8.3 からわかるように、学習者(JP2、JP3)と教科書（JP2_Textbook・JP3_Textbook)の Word Cloud は比較的近似しているが、学習者(JP2、JP3)と母語話者(CHN)の間には、全体的に大きな違いが見られる。

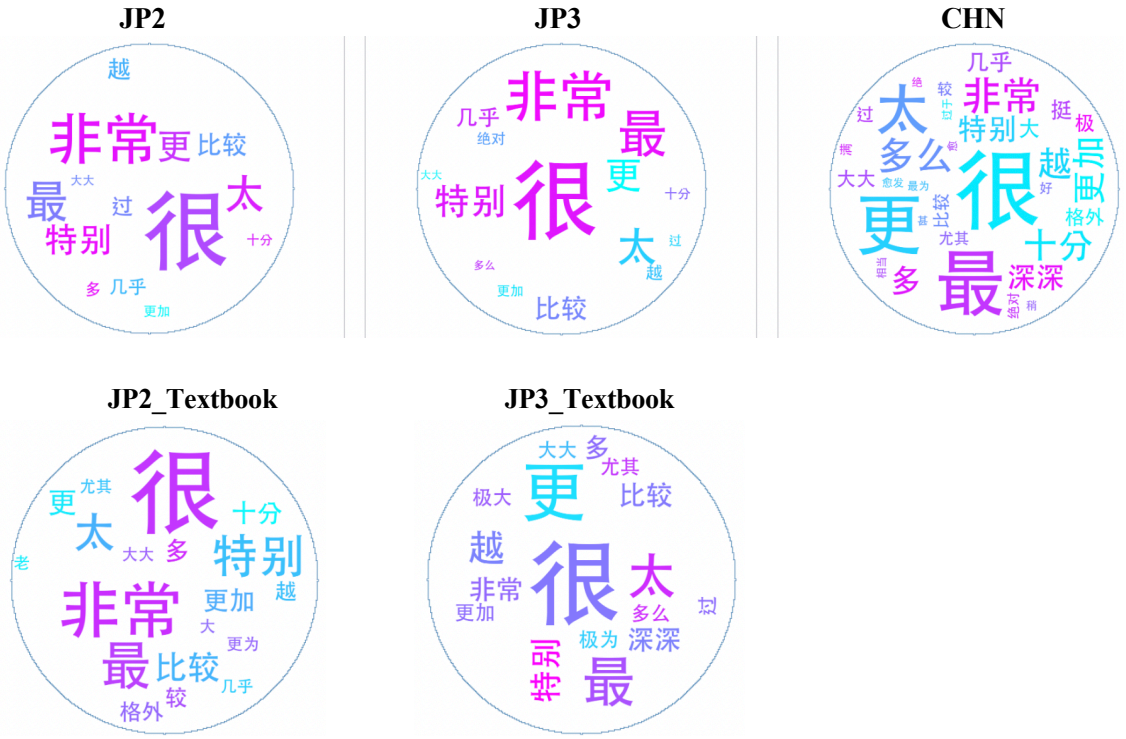

図 8.3　「程度を表す副詞」の Word cloud

表 8.7 は各サブコーパスの語彙別の「程度を表す副詞」の粗頻度と相対頻度を高い順に示した結果である。表 8.7 を見ると、「程度を表す副詞」の語彙の種類について、学習者の作文（JP2・JP3）コーパス、及び学習者の教科書（JP2_Textbook・JP3_Textbook）は、母語話者の作文（CHN）コーパスより著しく少ないことが確認された。

表 8.6 と表 8.7 を合わせてみると、学習者は作文を書く際、“很”（とても）、“非常”（非常に）などのような「程度を表す副詞」、“很多”（たくさん）、“我觉得”（*I think*）などのようなぼかし表現を頻繁に使っていることが分かった。また、学習者の作文における「程度を表す副詞」の使用頻度は母語話者の作文より著しく高いが、「程度を表す副詞」の語彙種類は非常に限られていることを明らかにした。これらの特徴や問題点は、日本語母語の負の転移と教科書の影響とも関係があると考えられる。なぜなら、「とても」、「非常に」などは、日本語の話し言葉でもよく使われている表現であり、“很”・“非常”などは、学習者の教科書にも頻繁に出現する語彙であるからである。また、学習者の教科書に現れた「程度を表す副詞」の語彙種類も非常に限られている。

表 8.7　語彙別の「程度を表す副詞」の頻度集計

JP2 (RTTR: 0.55)

No.	Words	Freq	R_Freq
	Tokens	3423	18.25
1	很	1726	9.20
2	非常	461	2.46
3	最	284	1.51
4	太	230	1.23
5	更	170	0.91
6	特别	165	0.88
7	比较	79	0.42
8	越	70	0.37
9	过	54	0.29
10	几乎	38	0.20
11	多	30	0.16
12	更加	19	0.10
13	十分	18	0.10
14	大大	16	0.09
15	大	11	0.06
16	尤其	8	0.04
17	多么	7	0.04
18	绝对	6	0.03
19	深深	5	0.03
20	相当	5	0.03
21	好	5	0.03
22	挺	3	0.02
23	较	3	0.02
24	蛮	2	0.01
25	何等	1	0.01
26	格外	1	0.01
27	极	1	0.01
28	稍微	1	0.01
29	极大	1	0.01
30	好不	1	0.01
31	老	1	0.01
32	愈发	1	0.01

JP3 (RTTR: 0.60)

No.	Words	Freq	R_Freq
	Tokens	2156	15.26
1	很	1089	7.71
2	非常	270	1.91
3	最	233	1.65
4	太	133	0.94
5	更	111	0.79
6	特别	104	0.74
7	比较	50	0.35
8	几乎	34	0.24
9	越	21	0.15
10	绝对	14	0.10
11	更加	12	0.08
12	过	12	0.08
13	十分	11	0.08
14	大大	9	0.06
15	多么	8	0.06
16	极	7	0.05
17	尤其	6	0.04
18	大	5	0.04
19	多	5	0.04
20	相当	5	0.04
21	挺	3	0.02
22	绝	3	0.02
23	万分	3	0.02
24	好	3	0.02
25	过于	2	0.01
26	特	1	0.01
27	深深	1	0.01
28	较	1	0.01

CHN (RTTR: 1.09)

No.	Words	Freq	R_Freq
	Tokens	4115	7.71
1	很	1013	1.90
2	最	654	1.23
3	更	558	1.05
4	太	355	0.67
5	非常	162	0.30
6	多么	150	0.28
7	越	125	0.23
8	十分	121	0.23
9	更加	114	0.21
10	多	94	0.18
11	特别	83	0.16
12	深深	78	0.15
13	几乎	47	0.09
14	挺	45	0.08
15	大	45	0.08
16	格外	32	0.06
17	极	31	0.06
18	大大	30	0.06
19	比较	27	0.05
20	过	25	0.05
21	尤其	21	0.04
22	较	20	0.04
23	好	18	0.03
24	满	16	0.03
25	绝对	15	0.03
26	最为	13	0.02
27	过于	13	0.02
28	愈发	12	0.02
29	相当	11	0.02
30	稍	10	0.02
31	绝	10	0.02
32	甚	10	0.02
33	愈	9	0.02
34	万分	8	0.01
35	极大	8	0.01
36	异常	8	0.01
37	稍微	8	0.01
38	极其	8	0.01
39	尤为	7	0.01
40	特	6	0.01
41	愈加	6	0.01
42	尤	5	0.01
43	何等	5	0.01
44	越发	5	0.01
45	全然	5	0.01

JP2_Textbook (RTTR: 1.57)

No.	Words	Freq	R_Freq
	Tokens	147	14.43
1	很	63	6.18
2	非常	20	1.96
3	最	14	1.37
4	特别	11	1.08
5	太	9	0.88
6	比较	5	0.49

JP3_Textbook (RTTR: 1.77)

No.	Words	Freq	R_Freq
	Tokens	92	13.75
1	很	37	5.53
2	更	15	2.24
3	最	10	1.49
4	太	8	1.20
5	越	4	0.60
6	特别	3	0.45

7	更	4	0.39	7	多	2	0.30	46	何其	4	0.01
8	更加	3	0.29	8	深深	2	0.30	47	颇	4	0.01
9	多	3	0.29	9	非常	2	0.30	48	更为	4	0.01
10	十分	3	0.29	10	比较	2	0.30	49	绝顶	4	0.01
11	格外	2	0.20	11	多么	1	0.15	50	老	4	0.01
12	越	2	0.20	12	更加	1	0.15	51	极为	4	0.01
13	较	2	0.20	13	过	1	0.15	52	略微	4	0.01
14	老	1	0.10	14	大大	1	0.15	53	极力	4	0.01
15	大大	1	0.10	15	尤其	1	0.15	54	蛮	3	0.01
16	大	1	0.10	16	极为	1	0.15	55	极端	3	0.01
17	更为	1	0.10	17	极大	1	0.15	56	分外	3	0.01
18	几乎	1	0.10					57	好不	3	0.01
19	尤其	1	0.10					58	极度	3	0.01
								59	颇为	3	0.01
								60	较为	3	0.01
								61	怪	2	0.00
								62	透顶	2	0.00
								63	至为	2	0.00
								64	大为	2	0.00
								65	不大	2	0.00
								66	忒	2	0.00
								67	倍加	1	0.00
								68	不胜	1	0.00
								69	更甚	1	0.00
								70	超	1	0.00

注：Tokens は総語数、「Freq」は粗頻度、「R_Freq」は 1,000 字あたりの相対頻度である。

(4) 「接続詞」の過剰使用について

例文（34）～例文（36）は「出来事の叙述」という作文ジャンから抽出した例文である。接続詞は、節や句をつなぐ論理的（logical）機能を持っており、情報量の多い談話に多く見られる（Biber 1988: 239）が、下記のような 3 つの抜粋に示したように、学習者が母語話者より、作文における接続詞が明らかに多く使用されていることが分かった。

例文：

(34) 从 这个 春天 ， 我 是 第二 年级 的 学生 。 <u>因为</u> 外国语系 的 校园 是 在 箕面 ， <u>所以</u> 我 要 去 哪儿 。 从 第二 年级 ， 我们 能 学习 第三 的 外语 。 <u>因为</u> 我 喜欢 韩国 电视 节目 ， <u>所以</u> 选 了 韩国语 。 那天 的 第五 课 就 是 韩国 课 。 高中 的 时候 ， 我 感 兴趣 韩国 。 从 前 我 独自 学习 韩国语 ， 我 会 说 一 点儿 。 <u>但是</u> 老师 教 给 我 韩国语 的 机会 是 初次 ， <u>所以</u> 我 很 期待 这个

课 。 **不过** 我 有 一个 不安 的 心情 。 **虽然** 几乎 我 的 朋 有 们 都 选 韩国语 课 ， **可是** 很 遗憾 ， 他们 都 是 别 老师 的 课 。

（JP2_出来事の叙述_022, Dimension 2 score: - 1.82, 接続詞: 25.55）

(35) 第三 天 ， 我们 去 台北 101 ， 这 座 在 台北 是 高楼大厦 。 我们 会 在 展望 台 观望 风景 ， 这 是 很好 的 。 **但是** ， 我们 一点儿 困 ， **因为** 我们 起床 早 ， **所以** 在 这儿 我们 小 睡 。 **然后** ， 我们 **再** 吃 刨冰 ， 喝 珍珠 奶茶 。 **然后** ， 我们 去 九 份 为了 看见 有名 的 地方 。 在 九 份 ， 有 很多 游客 ， **所以** 我 不 会 慢慢 儿 得 拍摄 。 **但是** ， 我 会 享受 了 那里 的 景色 。 **然后** ， 我们 去 按摩 。 在 这儿 ， 我 会 跟 姑姑 用 汉语 对话 。 我 的 朋友 不 会 说 汉语 ， **所以** 订购 或 对话 是 我 的 任务 。 我 用 汉语 说 ， 大概 懂得 。 **而且** ， 我 会 问 本地 人 路 ， 会 回答 他们 的 问题 。 通过 这个 之 行 ， 我 有 一点儿 自信 。

（JP3_出来事の叙述_056, Dimension 2 score: - 5.53, 接続詞: 51.59）

(36) 今天下午 ， 风和日丽 ， 天空 中 万里无云 。 天 显得 格外 明朗 。 我 仰望 蓝 天 ， 心想 ： 这 真 是 一个 放 风筝 的 好 天气 。 **于是** ， 我 **和** 妈妈 ， 哥哥 弟弟 妹妹 一起 去 仰韶公园 放 风筝 。 我们 兴高采烈 的 上 了 路 ， 一路 上 ， 我们 踏 遍 了 欢声笑语 ， 不一会 就 到达 了 目的地 。 此时 ， 我们 每个 人 的 脸上 都 挂 满 了 笑容 ， 每个 人 的 心中 都 洋溢 着 自信 **和** 骄傲 。 谁 都 想 让 自己 的 风筝 飞 得 最高 。 我 透过 车窗 外 看见 地上 的 枯草 冒 出 嫩 嫩 的 小 芽 儿 ， 柳树 也 发芽 了 ， 在 微风 的 吹动 下 像 小姑娘 舞动 的 头 发 。 走 到 公园 里 看见 放 风筝 的 人 可 真 多 啊 ！

（CHN_出来事の叙述_087, Dimension 2 score: 3.71, 接続詞: 18.65）

第 4 章で紹介した作文コーパスの検索機能を用い、JP2・JP3・CHN 作文コーパスにおける接続詞の出現頻度を集計した。図 8.4 は、JP2・JP3・CHN コーパスにおける接続詞の全体像を Word Cloud で可視化している。図 8.4 からわかるように、JP2 と JP3 には、“和”・“所以”・“可是”・“但是”などの単語が高頻度で現れている一方、CHN には “和”・“而”・“但”・“与”など単語が高頻度で出現している。

図 8.4　接続詞の Word cloud

表 8.8　語彙別の接続詞の頻度と割合

JP2 (RTTR: 1.23)				JP3 (RTTR: 1.45)				CHN (RTTR: 1.57)			
	語彙	頻度	割合		語彙	頻度	割合		語彙	頻度	割合
	Total	5027 (25.94)	100%		Total	4099 (27.95)	100%		Total	8421 (16.13)	100%
1	和	1277 (6.59)	25.4%	1	和	832 (5.67)	20.3%	1	和	1253 (2.40)	14.9%
2	所以	838 (4.32)	16.7%	2	所以	633 (4.32)	15.4%	2	而	987 (1.89)	11.7%
3	可是	393 (2.03)	7.8%	3	但是	452 (3.08)	11.0%	3	但	810 (1.55)	9.6%
4	但是	389 (2.01)	7.7%	4	可是	209 (1.43)	5.1%	4	与	392 (0.75)	4.7%
5	但	281 (1.45)	5.6%	5	而且	203 (1.38)	5.0%	5	因为	366 (0.70)	4.3%
6	因为	235 (1.21)	4.7%	6	因为	178 (1.21)	4.3%	6	可是	307 (0.59)	3.6%
7	不过	186 (0.96)	3.7%	7	但	170 (1.16)	4.1%	7	如果	268 (0.51)	3.2%
8	而且	184 (0.95)	3.7%	8	虽然	169 (1.15)	4.1%	8	虽然	249 (0.48)	3.0%
9	虽然	146 (0.75)	2.9%	9	如果	129 (0.88)	3.1%	9	所以	247 (0.47)	2.9%
10	然后	138 (0.71)	2.7%	10	不过	121 (0.83)	3.0%	10	但是	229 (0.44)	2.7%
11	而	94 (0.48)	1.9%	11	然后	109 (0.74)	2.7%	11	于是	203 (0.39)	2.4%
12	如果	93 (0.48)	1.9%	12	而	85 (0.58)	2.1%	12	然后	173 (0.33)	2.1%
13	因此	74 (0.38)	1.5%	13	即使	55 (0.38)	1.3%	13	并	142 (0.27)	1.7%
14	要是	45 (0.23)	0.9%	14	或者	49 (0.33)	1.2%	14	只要	135 (0.26)	1.6%
15	并	45 (0.23)	0.9%	15	或	46 (0.31)	1.1%	15	可	122 (0.23)	1.4%
16	另外	42 (0.22)	0.8%	16	要是	44 (0.30)	1.1%	16	只有	121 (0.23)	1.4%
17	或者	39 (0.20)	0.8%	17	首先	39 (0.27)	1.0%	17	还是	120 (0.23)	1.4%
18	或	37 (0.19)	0.7%	18	因此	36 (0.25)	0.9%	18	而是	107 (0.20)	1.3%
19	不仅	28 (0.14)	0.6%	19	不仅	30 (0.20)	0.7%	19	那么	106 (0.20)	1.3%
20	再说	26 (0.13)	0.5%	20	由于	30 (0.20)	0.7%	20	而且	104 (0.20)	1.2%
	…	…	…		…	…	…		…	…	…
87	故	1 (0.01)	0.0%	93	况且	1 (0.01)	0.0%	144	宁肯	1 (0.00)	0.0%

注：かっこ内に示したのは、1,000 字あたりの相対頻度である。

各サブコーパスの語彙別接続詞（主に上位 20 語）の頻度と割合を示した結果を表 8.8 に示す。表 8.8 における「頻度」列について、括弧外は粗頻度、括弧内は相対頻度の平均値である。表 8.8 からわかるように、JP2 と JP3 において接続詞の相対頻度（約 26 回、28 回）は CHN（約 16 回）の約 1.5 倍になるが、JP2 と JP3 において接続詞の語彙多様性（1.23、1.45）は CHN （1.57）より低いことが確認された。これにより、学習者の方が母語話者より全体的に接続詞の使用頻度が高く、上級になると接続詞の語彙多様性が増える傾向が見られるものの、学習者の語彙多様性は母語話者より低いことが分かった。

また、JP2 と JP3 における上位 12 位までの接続詞の順位は少し異なっているが、 接続詞の種類は同じであることが明らかである。これら 12 種の接続詞はそれぞれ、“和”・“所以”・“可是”・“但是”・“但”・“因为”・“不过”・“而且”・“虽然”・“然后”・“而”・“如果”である。それに対し、JP2・JP3 の上位 12 語 のうち 10 種の語彙が CHN の接続詞の上位 12 語に含まれているが、ほかの 2 語“与” ・“于是”は JP2・JP3 の上位 12 語に含まれていない。JP2・JP3 と CHN の上位 12 語は合計 14 種であり、これらの語彙の意味[85]の一部と難度の級別[86]は下記の通りである。

(1) “和”(L1)：と、および

(2) “所以”(L2)：だから、それで、そのために、それゆえに、そのために

(3) “可是”(L4)：しかし、だが、…であるが

(4) “但是”(L2)：しかし、だが、…であるけれども

(5) “但”(L2)：しかし、だが、…であるが

(6) “因为”(L2)：…のために、…なので、なぜなら…であるから

(7) “不过”(L4)：しかし、しかしながら、…であるけれど、ただ、だが

(8) “而且”(L3)：かつ、また

(9) “虽然”(L2)：それだけれど、それだが

(10) “然后”(L3)：それから、そして

---

[85] 接続詞の意味を整理する際、「中国語辞書 － Weblio 日中辞典・中日辞典」を参考にした。URL：https://cjjc.weblio.jp/〔最終アクセス日：2022 年 11 月 10 日〕

[86] L は Level の略であり、『新漢語水平考試(HSK)語彙』に基づき、語彙別の接続詞の難度を示している。例えば、 並列関係を表す“和”と“与”はそれぞれ HSK 1 級(最低級)と HSK 4 級(中級)の語彙、反転の意味を表す“可是”と“但是” はそれぞれ HSK 4 級(中級)と HSK 2 級(低級)の語彙であり、数字が大きくなるほど難度は高くなる。

(11) “而”(L4)：…であるがしかし…、かつ、また

(12) “如果”(L3)：万一、若しかして、もし～

(13) “与”(L4)：…と…、…そして…、…および…

(14) “于是”(L4)：それで、そこで

JP2 と JP3 には、上位 12 語の接続詞が総頻度の約 84.6%と 80.3%を占めているのに対し、CHN の上位 12 語の接続詞は全体の約 65.1%しか占めていない。興味深いのは、使用頻度の最も高い 4 つの接続詞である。JP2 と JP3 における上位 4 語の接続詞の種類が同じであり、“和”(L1)・“所以”(L2)・“可是”(L4)・“但是”(L2)の総使用率が接続詞総頻度の約 57.6%と 51.9%を占めているのに対し、CHN では“和”(L1)・“而”(L4)・“但”(L2)・“与”(L4)の順で総頻度の約 40.9%を占めているにすぎない。

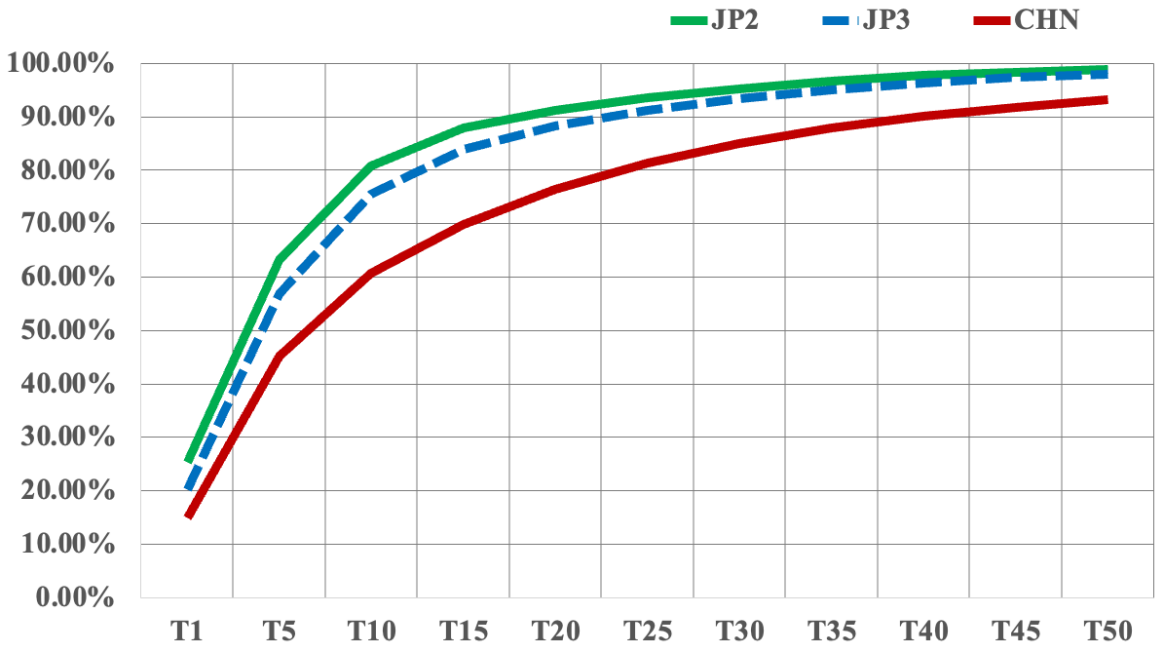

図 8.5 接続詞：累積構成比の変化

また、文章における語彙は、一般的に、少数の高頻度語と多数の低頻度語から構成されると石川（2012: 152）は述べている。学習者（JP2、JP3）と母語話者（CHN）の作文コーパスに出現している高頻度接続詞が全体に占める比率は、中国語レベルの低い学習者作文コーパスの方が高いという仮説を立てることができる。この仮説を検証するため、

各コーパスの接続詞において、頻度別の構成比を調べ、頻度が極めて高いとみなされる上位 1 語（T1、T は Top の略）、5 語（T5）、 10 語（T10）、中低位語とみなされる 15 語（T15)、20 語（T20）、25 語（T25）、 30 語（T30）、35 語（T35）、40 語（T40）、45 語（T45）、50 語（T50）の累計構成比を求めた。接続詞における頻度別構成分析の結果は図 8.5 の通りである。

図 8.5 を観察すると、JP2 と JP3 において、接続詞の累計構成比の変化が近似している。また、T1 語からの構成比率は、JP2 と JP3 が一貫して CHN を上回ることが確認され、T10 語以上になると差は次第に縮小する。JP2 と JP3 には、高頻度とみなされる上位 10 語の接続詞が総頻度の約 80.9%と 75.5%を占めているのに対し、CHN の上位 10 語の接続詞は全体の約 60.7%しか占めていない。したがって、学習者の作文で使われる接続詞の高頻度語比率が高いことが確認された。

接続詞は「前後の文脈の関係を表すもの」（石黒 2008: 44）として、文、節、句、語などを繋ぐ機能を持っており、文章を作る上で非常に重要な役割を果たしている。接続詞の過剰使用や過少使用は、必ずしも誤用を意味するものではない（Heino, 2010; Narita *et al.* 2004）あるものの、接続詞を適切に使わないと、文章のまとまりや読みやすさが損なわれてしまう（Larsen-Walker 2017）場合は少なくない。「中国語の複文の構造では、接続詞類がなくても文を繋ぐことが可能」（王 2009）であると述べられているように、母語話者の作文（例えば、例文 36）では、接続詞を使用しなくとも文脈の前後関係を理解できる。それに対して、学習者の作文では、接続表現が必要ではない文脈においても接続詞が出現しており、似たような意味を持つ他の接続詞に置き換えることなく、同じ接続詞を何度も重複に使用するケースが散見される。例えば、上記の例文（34）～（35）において、 因果関係を表す接続詞“<u>因为</u>～<u>所以</u>”（訳：…なので～だから…）、反転の意味を示 “<u>但是</u>”（訳：しかし）、時間的な前後関係を表す“<u>然后</u>”（訳：その後、それから）は何回も使用されている。この現象は、学習者の語彙の乏しさを反映しているとも言える。

学習者の作文に反映された接続詞語彙の乏しさは、教科書に現れた接続詞の影響に関係があると考えられる。表 8.9 は作文教科書の例文における接続詞を高い順に示している。表 8.9 から分かるように、教科書の作文例文に現れた接続詞の種類は、表 8.8 に示した中国語母語話者作文における接続詞の種類より著しく少ない。学習者が教科書で触

れる接続詞の種類は少なく、当然ながら作文における接続詞の語彙多様性に欠けると推察される。

表 8.9　教科書における語彙別の接続詞の頻度（上位 20 語）

JP2_Textbook				JP3_Textbook			
No.	**Words**	**Freq**	**R_Freq**	**No.**	**Words**	**Freq**	**R_Freq**
	Tokens	211	20.71		Tokens	147	21.96
1	和	36	3.53	1	和	21	3.14
2	但	19	1.86	2	但	14	2.09
3	所以	18	1.77	3	但是	11	1.64
4	而	11	1.08	4	所以	8	1.20
5	不过	10	0.98	5	而	7	1.05
6	而且	9	0.88	6	如果	7	1.05
7	但是	9	0.88	7	而且	6	0.90
8	因为	8	0.79	8	虽然	5	0.75
9	虽然	7	0.69	9	不仅	5	0.75
10	因此	7	0.69	10	因为	4	0.60
11	可是	6	0.59	11	因此	4	0.60
12	可	5	0.49	12	总而言之	4	0.60
13	如果	5	0.49	13	那么	4	0.60
14	于是	4	0.39	14	可是	4	0.60
15	既	4	0.39	15	而是	3	0.45
16	并	3	0.29	16	然后	3	0.45
17	不仅	3	0.29	17	或	3	0.45
18	不管	3	0.29	18	甚至	3	0.45
19	再说	3	0.29	19	可见	3	0.45
20	而是	3	0.29	20	只要	2	0.30
…	…	…	…	…	…	…	…
48	不只	1	0.10	38	尽管	1	0.15

注：Tokens は総語数、「Freq」は粗頻度、「R_Freq」は 1,000 字あたりの相対頻度である。

8.2.2 言語項目の過少使用について

8.1 節に述べたように、作文における「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、学習者が母語話者より全体的に著しく低い。また、「人物描写」という作

文ジャンルでは、学習者の方が、「動作動詞」・「方向動詞」・「高頻度動詞」の使用が明らかに不十分である問題点があることも確認した。

(1)「高難度語彙」・「非 HSK 語彙」の過少使用について

「高難度語彙」・「非 HSK 語彙」・「低難度語彙」は中国語 HSK 試験に関わる語彙リスト[87]を参考にして定めた言語項目である。表 8.10 は、母語話者（CHN）と学習者（JP2、JP3）サブコーパスに加え、中級中国語作文授業と上級中国語作文授業で使用された作文教科書における作文例文（それぞれ 17 編と 12 編）に含まれている HSK 語彙と非 HSK 語彙の割合を示している。

表 8.10　作文教科書の例文と作文コーパスにおける HSK 語彙の割合

Subcorpus	低難度語彙（HSK1-2 級）	中難度語彙（HSK3-4 級）	高難度語彙（HSK5-6 級）	非 HSK 語彙
CHN	45.49%	15.79%	9.75%	28.97%
JP2	56.02%	15.08%	5.45%	23.46%
JP3	51.54%	16.94%	7.71%	23.80%
Textbook_JP2	53.02%	16.63%	6.27%	24.08%
Textbook_JP3	46.00%	19.95%	9.51%	24.54%

注：Textbook_JP2 は中級作文授業の教科書であり、Textbook_JP3 は上級作文授業の教科書である。

表 8.10 から分かるように、L2 中国語作文教材は主に HSK 語彙を中心に編纂されたものだといえる。その理由は、中級と上級中国語作文教科書の例文において、HSK 語彙が総語数の約 76%を占めているのに対し、非 HSK 語彙が全体の約 24%しか占めていない。また、学習者（JP2、JP3）の作文コーパスと中級中国語作文教科書の例文のおける低難度語彙の割合は、総語数の 50%を超えている。さらに、非 HSK 語彙の割合については、学習者（JP2、JP3）の作文コーパスは中級・上級中国語作文教科書と近似しており、と

[87] 第 2 章に述べた通り、本研究の言語項目を特定した際、「HSK 語彙リスト（5000 語）」を低難度 （HSK1-2 級）、中難度 （HSK3-4 級）、高難度（HSK5-6 級）というような 3 段階の語彙リストに分けた。非 HSK 語彙は「HSK 語彙リスト（5000 語）」に収録されていない語彙を指すことである。また、第 5 章で解明したように、これらの 4 つの言語項目において、非 HSK 語彙・高難度語彙・低難度語彙のような 3 つの言語項目は次元 1 に強い影響を与えていることが分かった。

もに約 24%であるものの、母語話者（約 29%）より低いことを明らかにした。学習者は、目にする機会が多い低難度語彙を過剰に使用しているが、教科書で頻繁に使用されていない非 HSK 語彙や高難度語彙は、学習者が目にする機会が乏しく、学習者の産出した語彙にあまり含まれていないと推察される。そのため、非 HSK 語彙や高難度語彙の使用について、学習者は語彙力の高い中国語母語話者の程度に達していないと考えられる。

(2) 「時態助詞“着”」の過少使用について

作文テキストの抜粋を合わせてみると、母語話者の作文ジャンルでは、「時態助詞“着”」を用いることで、状態状況表現が多く使用されているのに対し、学習者作文ジャンルの多くでは、状態状況表現があまり用いられていない傾向が見られた。

例文：

(37) 我 爸爸 先 开始 放 风筝 ， 给 我 示范 了 一 遍 ， 并 说 出 放 风筝 的 一些 要领 。 看 <u>着</u> 爸爸 放 风筝 那 轻松 的 样子 ， 我 也 想 试试 ， 就 吵 <u>着</u> 爸爸 让 他 停 下来 ， 让 我 自己 来 放 风筝 。 起初 ， 我 也 像 爸爸 一样 ， 右手 拿 <u>着</u> 风筝 ， 左手 拿 <u>着</u> 风筝 线 ， 学 <u>着</u> 爸爸 的 动作 ， 迎 <u>着</u> 风向 ， 快速 奔跑 ， 把 风筝 放 向 上空 。 然后 ， 一边 跑 一边 放 线 ， 风筝 很快 就 飞 起来 了 。 我 慢慢 地 放 线 ， 风筝 越 飞 越 高 。

（CHN_出来事の叙述_086, Dimension 2 score: 5.50, 時態助詞“着”: 12.05）

(38) 这 是 对 我 第一 次 去 冲绳 的 经验 。 第一 天 ， 我们 去 了 “ フサキビーチ ” ， 这里 是 一 家 饭馆 的 私有 海滩 ， 所以 人 不 多 。 我 跟 她 一起 游泳 、 和 果汁 ， 过 得 很 开心 。 我们 去 乘 水上 摩托艇 ， 我们 坐 游泳 圈 ， 然后 水上 摩托艇 啦 游泳 圈 。 这个 游戏 很 好 意思 。 <u>我 和 她 认真 的 拿 扶手</u> ， 因此 手 解放 的 话 ， 要 海里 跳 下去 。 晚上 吃 了 寿司 ， 都 是 冲绳 的 海鲜 ， 味道 很 好 。 第二 天 ， 我们 座 船 去 竹富岛 ， 要 十五 分钟 左右 就 到 了 。

（JP3_出来事の叙述_058, Dimension 2 score: - 7.43, 時態助詞“着”: 0.00）

例えば、例文（37）は、“看 着”・“吵 着”・“拿 着”・“学 着”・“迎 着”のような「動詞＋時態助詞“着”」というパターンを使用することで、父親と凧揚げをする出来事の臨場感を鮮やかに描写することができた。

それに対して、学習者の作文では、「時態助詞“着”」が必要である文脈においても「時態助詞“着”」が使用していないケースが見られる。例えば、例文（38）に示したように、「私と彼女は、モーターボートのハンドルをしっかり握って……」のような日本語の意味を表す場合、学習者は“我 和 她 认真 的 拿 扶手……” というセンテンスを書いている。この短いセンテンスには 2 種類の問題がある。1 つは語彙選択の問題で、もう 1 つは「時態助詞“着”」を使用していないことである。まず、語彙選択の問題について、「……をしっかり握って」における「しっかり」の意味を表す最も自然な中国語の表現は“认真 的”（訳：真面目に）ではなく、“紧紧 地”である。また、モーターボートの「ハンドル」を表す中国語の語彙について、“扶手”（訳：手すり）があまり自然ではなく、“车把”や“把手”が最も自然な表現であるだろう。「時態助詞“着”」の使用に関しては、「握って」という表現を中国語で言うと、「握 着」が最も自然な表現になると考えられる。しかし、学習者はこの意味を伝えるために“拿”（訳：持つ、取る）という語彙しか使っていない。

日本語：私と彼女は、（モーターボートの）ハンドルをしっかり握って……

学習者：我 和 她 认真 的 拿 扶手……

自然な表現：我 和 她 紧紧 地 握 着 （摩托艇 的）车把……

注：かっこ内の内容は省略可

母語話者はジャンルに応じて、適切な状態状況表現を使うことができるが、適切な語彙の選択をまだ習得できていない学習者は、ジャンルに応じた適切な状態状況表現を習得する段階には至っていないことが推察される。

(3) 「時間を表す副詞」の過少使用について

例文(39) ～ (41) は、CHN・JP3・JP2 のサブコーパスから抽出された人物描写の抜粋である。下記の例文において、「高頻度副詞」には下線が引かれており（例：“很”、“真”，“还”）、「時間を表す副詞」は下線・イタリック体・傍点（例：“*有时*”、“*总是*”，“*经常*”）で表示されている。

例文：

(39) 小阳 是 我 的 好 朋友 ， 整天 乐呵呵 的 ， *有时候* 比 男生 还 顽皮 ， *有时* 又 比 女生 还 安静 ， 一 双 水灵灵 的 大 眼睛 *总是* 滴溜溜 的 转 着 ， 似乎 有 出 不 完 的 鬼主意 。那 是 一个 阳光 明媚 的 早晨 。 吃 完 早饭 ， 我 发现 小阳 *不断* 的 绕 着 有 玫瑰花 的 花坛 转 ， 她 左 瞧瞧 右 看看 ， 双手 *不断* 扒开 草丛 ， 似乎 在 寻找 什么 。 我 好奇 的 跑 到 她 的 身边 ， 问道 ： “ 小阳 ， 你 在 干 什么 呀 ？ ” 小阳 神秘 的 说 ： “ 这 是 秘密 ， 你 等 一会儿 就 知道 了 。 ” 于是 我 便 *一直* 跟 着 小阳 寻找 ， 我 *一直* 跟 着 她 ， 她 装作 很 生气 的 跑 了 ， 但 等 我 走 后 ， 小阳 又 *继续* 找 了 起来 。 放 学 了 ， 回家 路上 ， 小阳 满脸 笑容 地 问 我 ： “ 妮妮 ， 你 猜 我 在 花坛 里 找到 了 什么 ？ ” 我 无知 的 摇 了 摇 头 ， 她 又 说 ： “ 你 把 手 拿 出 来 就 知道 了 。 ” 我 半信半疑 的 将 手 伸 了 过去 ， 小阳 *立马* 将 一个 玻璃 罐子 倒扣 在 我 手上 ， 当 她 把 罐子 拿 起来 的 一瞬间 ， 我 惊 呆 了 ， 竟然 是 一 条 蚯蚓 ！ 我 吓 得 狂 甩 不 止 ， 那 小阳 却 大 笑 着 跑 开 了 。 唉 ， 真 是 个 “ 捣蛋 鬼 ” 。

（CHN_人物描写_010, Dimension 3 score: 1.70, 高頻度副詞: 56.68, 時間を表す副詞: 34.41）

(40) 我 有 一个 十五 年 来 的 挚友 。 她 今年 二十二 岁 。 她 中等 的 身材 ， 瓜子 脸 ， 眼睛 很 美丽 。 她 和 我 *曾* 上 了 同 一 所 小学 、 初中 和 高中 。 虽 然 大学 不同 ， 但是 我们 *经常* 见面 一起 玩儿 。 她 开朗 的 性格 好像 一 朵 向日葵 一样 。 她 *总是* 带 着 笑容 ， 说 很多 有趣 的 话 ， 引 大家 发笑 。 她 华丽 的 外表 好像 一 朵 玫瑰 一样 。 她 每天 锻炼 身体 ， 研究 新 的 化妆 艺术 ， 购买 漂亮 的 衣服 。 就是 说 她 是 一个 内外 兼备 的 美丽 人 。 我

认识 她 是 在 小学 一 年级 的 时候 。 当时 她 眼睛 大大 的 ， 皮肤 白白 的 ， 像 个 洋娃娃 似的 。 由于 我 的 座位 离 他 的 很 近 ， 所以 我们 的 关系 *马上* 就 变得 很 好 了 。 在 学校 休息 的 时间 ， 我们 在 操场 的 一个 角落 里 *一直* 聊天 。 我们 谈 了 关于 昨天 看 的 电视 节目 ， 最近 喜欢 的 男孩 和 将来 的 梦想 等等 。

（JP3_人物描写_062, Dimension 3 score: - 1.70, 高頻度副詞: 34.87, 時間を表す副詞: 15.34）

(41)有 这么 一 位 帅哥 ， 他 长 得 非常 帅 。 干松 的 头发 ， 光润 的 脸 ， 下 有 一 双 茶色 的 眼睛 ， 嘿嘿 ， 是 不 是 挺 帅 呢 ？ 他 很 擅长 唱歌 。 他 参加 大阪大学 的 合唱团 。 他 热心 练习 唱歌 ， 所以 他 的 进步 很 快 。 因为 他 的 声音 真 漂亮 ， 社团 成员 被 他 的 歌声 迷倒 。 有 一 天 他 独唱 的 时候 ， 很多 女 成员 夸奖 他 。 别 的 男 成员 都 羡慕 死 他 了 。 他 也 擅长 学习 。 他 是 外语 学院 中文 系 的 学生 。 他 用功 学习 。 他 说 汉语 说 得 很 流利 。 上 次 在 大学 的 食堂 吃 午饭 ， 跟 中国 留学生 流利 的 聊天 儿 。 他 背 多 汉语 单词 。 他 每天 给 朋友 教 汉语 ， 所以 他 很 受 欢迎 。 他 想 去 北京 留学 吃 烤鸭 。 大学 毕业 以后 ， 他 想 回 故乡 当 公务员 。他 还有 许多 优点 。 可 美中不足 的 就是 会话 水平 低 。 比如 合唱 练习 的 时候 ， 虽然 大家 享受 和 朋友 们 聊天儿 的 乐趣 ， 但是 他 *常常* 呆呆 地 站 着 。 他 就是 不 善于 跟 人 说话 了 。

（JP2_人物描写_007, Dimension 3 score: - 1.62, 高頻度副詞: 44.59, 時間を表す副詞接: 9.55）

例文(39)では、友達である小陽（“小阳”：陽さん）という登場人物を描写する際、ミミズ捕りの出来事の経緯を記述するために、“*有时*”（訳：時には）、“*总是*”（訳：いつも）、“*不断*”（訳：絶えずに）、“*立马*”（訳：すぐに）など、高頻度副詞や「時間を表す副詞」を豊かに用いることで、小陽のいたずら好きな性格を動画のように生き生きと表現している。

それに対して、例文(40)と例文(41)に示すように、学習者、特に中級レベル(JP2)の学習者が作文で用いる副詞や「時間を表す副詞」が中国語母語話者に比べて著しく少ない。特に例文(41)では、短いセンテンスが積み重なり、副詞修飾語があまり使われてい

ない。登場人物の特徴や個性を鮮やかに際立たせるという点では、例文(39)の描写が動画のようだとすれば、例文(41)の描写は写真を見せるようなものだと言えるだろう。原因の一つは、学習者の中国語言語能力の不足と関係があると思われる。上級（JP3）になると、作文で用いる副詞や「時間を表す副詞」の量が増える傾向が見られるからである。

表 8.11　語彙別の「時間を表す副詞」の頻度集計（上位 10 語）

JP2				JP3				CHN			
No	Words	Freq	R_Freq	No	Words	Freq	R_Freq	No.	Words	Freq	R_Fre
	Tokens	2,894	15.43		Tokens	2,129	15.07		Token	13,535	25.36
1	都	580	3.09	1	都	393	2.78	1	就	2461	4.61
2	就	428	2.28	2	就	354	2.51	2	都	1874	3.51
3	还	210	1.12	3	还	129	0.91	3	又	1148	2.15
4	已经	200	1.07	4	已经	108	0.76	4	还	1037	1.94
5	又	151	0.81	5	一直	108	0.76	5	才	583	1.09
6	常常	140	0.75	6	常常	96	0.68	6	再	496	0.93
7	一直	121	0.65	7	又	95	0.67	7	便	376	0.70
8	经常	104	0.55	8	才	74	0.52	8	总是	325	0.61
9	再	94	0.50	9	再	71	0.50	9	已经	321	0.60
10	才	81	0.43	10	经常	69	0.49	10	还是	235	0.44
…	…	…	…	…	…	…	…	…	…	…	…
96	早就	1	0.01	87	常年	1	0.01	159	顿然	1	0.00

JP2_Textbook				JP3_Textbook			
No	Words	Freq	R_Freq	No	Words	Freq	R_Freq
	Tokens	250	24.54		Tokens	154	23.01
1	都	62	6.08	1	就	36	5.38
2	就	38	3.73	2	都	35	5.23
3	还	36	3.53	3	还	16	2.39
4	才	11	1.08	4	又	9	1.34
5	已经	10	0.98	5	已经	6	0.90
6	一直	10	0.98	6	原来	5	0.75
7	又	8	0.79	7	刚	4	0.60
8	正	6	0.59	8	才	4	0.60
9	常常	6	0.59	9	不断	3	0.45
10	总	5	0.49	10	终于	3	0.45
…	…	…	…	…	…	…	…
49	始终	1	0.10	35	常常	1	0.15

注：Tokens は総語数、「Freq」は粗頻度、「R_Freq」は 1,000 字あたりの相対頻度である。

また、表 8.11 からわかるように、学習者の作文における「時間を表す副詞」を過少に使用している原因の一つは、教科書の例文に現れる「時間を表す副詞」の不足に関係があると考えられる。学習者が使用した教科書において、「時間を表す副詞」の語彙の種類は非常に限られている。中国語習得の段階では、知識や語彙のインプットが不足すると、該当する語彙は学習者の産出した言葉にあまり含まれないと考えられる。

本研究では、学習者が使用した作文教科書における人物描写（中国語：写人记叙文）の例文（それぞれ 1 篇）を下記のように引用した。作文例文における副詞や「時間を表す副詞」を調べた結果、JP2（中級作文授業）の教科書において、副詞の出現頻度や異なり語数が少ないが、JP3（上級作文授業）になると、教科書における副詞の使用頻度や副詞語彙の豊富さが増える傾向が見られた。したがって、JP2（中級学習者）の作文で副詞の使用が少ないこと、JP3（上級学習者）作文で副詞の使用が増えることも、教科書の影響と関係があると思われる。

JP2 教科書における人物描写の例文：《我的哥哥》「私のお兄さん」

我 的 哥哥 身材 高大 ， 身高 有 一 米 九 以上 。 他 不 是 很 英俊 ， 但 是 因为 他 个子 高 ， 所以 引人注目 。 他 视力 很 不好 所以 不得不 戴 眼镜 。 我 觉得 他 的 外貌 是 典型 的 日本 人 。 不过 我 和 哥哥 不 怎么 像 。

哥哥 是 一个 很 小心谨慎 的 人 ， 如果 有 了 想 做 的 事情 ， 他 会 先 查 一 查 ， 考虑 以后 才 开始 行动 。 所以 他 很少 表现 出 着急 的 样子 。

不过 我 见 过 一 次 他 着急 的 样子 ， 是 知道 妻子 怀孕 的 消息 后 。 那时 他 既 高兴 ， 又 担心 。 也 是 ， 别 的 事情 都 可以 准备 一 下 ， 考虑 一下 ， 这个 事情 恐怕 很 难 提前 准备 什么 了 。 他 在 电话 里 告诉 我 这 件 事情 的 时候 ， 开始 什么 都 不 说 。 我 觉得 是 因为 他 不 知道 怎么 告诉 我 ， 后来 是 在 嫂子 的 帮助 下 才 说 了 出来 。 那时 他 的 声音 充满 幸福 ， 充满 活力 。 我 从来 没 听 过 他 那么 高兴 的 声音 呢 ！

我 相信 ， 他 能 成为 一个 既 有 责任感 又 顾家 的 好 爸爸 。

注：亓华・贾丹丹(2013) の『汉语・纵横 写作课本(上)』p.96 に掲載されている例文より引用

JP3 教科書における人物描写の例文：《善良的妈妈》「優しいお母さん」

妈妈，一个多么让人感到温暖的词语。每当我想起我的妈妈，一股暖流就涌上心头。

她中等个子，身体很结实，虽然已经四十多岁了，可是见过我妈妈的人都说："你妈妈的眼睛真美。"是的，我的妈妈或许不是很漂亮，但她却有着一双明亮的大眼睛，给人的印象很深。俗话说："眼睛是心灵的窗户。"看到这双明亮的眼睛，你就知道我的妈妈有一颗善良的心。

妈妈很善良。记得小时候，有一天，我们一家去动物园玩儿。其中有一个娱乐项目叫做"山羊拉车"。一头怀孕的山羊拉着一辆小车，小孩坐在车上玩儿，一元钱一次。那天正好是"六一"儿童节，动物园的生意好得不得了。老板一会儿工夫就赚了不少钱，但是那只可怜的山羊，快要做妈妈了，却不得不被迫拉着沉重的车，一圈一圈地给主人赚钱。不懂事的我也叫嚷着要去坐"山羊车"，回头却看见妈妈的眼睛里噙满了泪花。她轻轻地说："我们不坐，山羊要做妈妈了，不能累坏了她。"说完，我的妈妈走到老板跟前，大概是和他商量什么，只记得妈妈从兜里掏出了一些钱，山羊的主人很不情愿地结束了这个游戏。也许是因为妈妈的诚恳打动了他。

回家的路上，妈妈说："山羊和我们一样，有妈妈，有宝宝，我们都要爱护她。"现在我已经长大了，但这句简单的话却深深刻在我的心里。这就是妈妈给我人生的最大财富——一颗金子般善良的心。

注：李汎（2009）の『汉语综合写作教程』p.53 に掲載されている例文より引用

(4)「平均文長」の短さと「語彙多様性」の低さについて

L2 ライティング研究の分野では、「平均文長」と「語彙多様性」が学習者の言語能力を反映する重要な指標となっている。学習者の言語能力が高いほど、作文の平均文長が長くなり、語彙の多様性が高くなるとよく言われている（例、王 2015; Yang *et al.* 2015; Ai & Lu 2013; 呉 2016; 呉 2019）。本研究では、学習者の作文例文を組み合わせながら考察した結果、母語話者に比べ、学習者の作文、特に JP2（中級学習者）の作文は、セン

テンスの長さが短く、短いセンテンスが積み重なっており、語彙の繰り返しが多いことが確認された。

この問題点について、学習者の作文における平均文長を長くするために短文を統合して長文を書く訓練が必要であり、また語彙多様性を高めるために同義語表現を使う練習を強化する必要があると考えられる。

(5) 「人物描写」作文の物語性の乏しさについて

人物描写の作文を書く際、学習者の作文は、単に人物の外見の特徴と所属などに焦点を当てることが好まれ、具体的な出来事を詳細に叙述することで人物の特徴や個性を引き立てることが少なかったため、「動作動詞」（傍点）・「方向動詞」（*イタリック体*）・「高頻度動詞」（下線）の使用が、母語話者の作文に比べて低い。

---

(28) 我 家里 有 一个 淘气 无敌 的 弟弟 ， 名叫 小球球 。 小球球 有 一 双 炯炯有神 的 眼睛 ， 眉毛 像 刷子 一样 ， 一 张 能言善辩 ， 爱 拍 “ 马屁 ” 的 小 嘴 巴 ， 一 只 小巧玲珑 的 鼻子 ， 脸蛋 粉红 粉红 的 ， 像 一个 又 大 又 红 的 苹果 。 只要 你 口袋 里面 有 着 钱 ， “ 嗖嗖 ” 地 一下子 ， 他 把 钱 全 都 变成 了 糖果 和 食物 。 记得 有 一 次 ， 妈妈 带 我 和 小球球 *去* 超市 购物 ， 小球球 看见 了 一 把 玩具 小 水枪 ， 边 摇 妈妈 的 手 边 说 ： “ 好 妈妈 ， 你 看 ， 那 把 小 水枪 如果 挂 在 我 身上 肯定 又 帅 又 好看 ， 长大 以后 保护 妈妈 ， 球球 想 要 那个 玩具 ， 你 就 如 了 我 的 心愿 吧 ！ ” 妈妈 听 了 他 的 甜言蜜语 ， 给 小球球 买 了 *下来* 。 *回家* 后 ， 小球球 就 开始 玩弄 *起* 这 把 水枪 。 妈妈 还 想 听 听 那 甜言蜜语 ， 就 走 *去* 说 ： “ 小球球 ， 妈妈 都 给 你 买 了 ， 你 是 不 是 …… ” 妈妈 话 还 没 说 完 ， 只 听见 “ 哇 ” 的 一 声 ， 我 立马 从 房间 跑 出来 ， 我 的 乖乖 ， 小球球 又 拿 小 水枪 打 了 妈妈 一 枪 ， 真 淘气 。

（CHN_人物描写_008, Dimension 4 score: 3.62, 動作動詞: 74.75, 方向動詞: 20.20, 高頻度動詞: 127.27）

(42) 我 介绍 一个 朋友 ， 跟 她 社团 活动 和 学系 都 相同 。 她 个子 很 矮 ， 柔 长 的 头发 ， 白白 的 皮肤 ， 大大 的 眼睛 ， 非常 可爱 。 第一 次 见面 是

入学 典礼 后 的 说明会 。 我 **觉得** 她 很 可爱 ， **想** 和 他 **说话** ， 但 我 **没有** 勇气 。 她 好像 **交** 了 几 个 朋友 ， 我 **觉得** 已经 来不及 了 。 然而 ， 几 天 后 ， 我 犹豫 **加入** 美式 足球 社团 时 ， 她 已经 **决定** **加入** 美式 足球队 。 我 也 被 美式 足球队 所 **吸引** ， 可是 还 犹豫 。 原因 是 因为 我 家 **离** 学校 很 远 ， 我 **担心** **没有** 时间 学业 和 社团 活动 兼顾 。 但是 ， 她 也 老家 **离** 学校 很 远 ， 所以 我 鼓 ***起*** 勇气 **决定** **加入** 美式 足球队 。 可以 毫不 夸张 的 **说** ， 我 **加入** 的 原因 中 ， 她 约 **占** 60% 。 我们 **花** 了 很 长 的 时间 从 一起 ***过*** 了 学生 生活 。 由于 学系 和 社团 活动 都 是 一样 的 ， 我们 每天 **见面** ， **有** 很 多 共同 话题 。 我 很 高兴 ， 因为 没 **想到** 在 大学 里 能 **交** **到** 这样 的 朋友 。

（JP3_人物描写_056, Dimension 4 score: -2.33, 動作動詞: 40.15, 方向動詞: 3.65, 高頻度動詞: 93.07）

(43) **有** 一个 很 好看 的 女人 ， 她 为什么 好看 ？ 她 **有** 乌黑 的 头发 和 整洁 的 眉毛 。 他 鼻子 **像** 匹诺曹 一样 高 ， 个子 也 高 。 而且 ， 她 的 笑颜 好像 **开花** 很 可爱 。 以前 她 **住** 在 乡下 ， 为了 上 大学 ， 她 从 去年 **开始** **住** 在 大阪 。 大阪 是 一个 很 大 的 城市 ， 她 **发现** 了 在 那儿 大多数 人 **打扮** 得 漂漂亮亮 。 她 **想** ： “ 我 也 **穿** 得 漂亮 。 ” ， 最近 着迷 用 网路 **买** 各种各样 的 衣服 。 再说 ， 她 **开** 一个 人 **住** 已经 **有** 一 年 多 了 ， 可 是 因为 她 想 舒适 地 ***过*** 天 天 ， 所以 至今 **收集** 可爱 的 杂货 。 比如说 ， 干燥 花 、 蜡烛 、 景色 的 照片 等等 。 她 的 房间 不 太 大 ， 她 **买** 的 东西 越 多 房间 越 狭小 ， 可 是 她 爱 干净 ， 房间 里 总是 **整理** 得 井井有条 。他 很 温柔 ， 不 是 当机立断 的 人 。 **看** 老人 **搬** 大 包 的 时候 ， 她 **想** **帮助** 他 ， 可是 只 **看** 这个 样子 。 在 电车 上 ， 她 **想** 和 老人 或者 孕妇 **换** 座位 ， 可 是 很 难 出口 ， 终于 **到达** 目的地 。 这些 都 是 对 日本 人 难能可贵 ， 是 不 是 ？ 她 是 典型 的 日本 人 。

（JP2_人物描写_022, Dimension 4 score: -3.23, 動作動詞: 25.56,方向動詞: 3.19, 高頻度動詞: 92.65）

第 7 章に示した例文（28）と本章で示した例文（42）～（43）を合わせてみると、人物描写という作文を書く際、母語話者はキャラクターの外見を描くだけでなく、特定の出来事を通してキャラクターの特徴や性格を際立たせている。一方、学習者は、主にキ

ャラクターの外見や所属などの特徴に焦点を当てることが好まれ、具体的な出来事を詳細に叙述することでキャラクターの特徴を引き立てることが得意ではないようである。具体的な出来事の詳細を叙述するのであれば、「動作動詞」や「方向動詞」の出現頻度は多くなるであろう。しかし、学習者は具体的な出来事を語りながら人物を描写することが苦手であるため、作文における「動作動詞」や「方向動詞」の頻度が母語話者より低くなってしまったと考えられる。

8.2.3 中国語作文指導への提言

上記の分析に基づき、学習者の産出した書き言葉における言語特徴の過剰使用や過少使用の原因は、主に母語である日本語の干渉、教科書に表れた語彙の影響などに帰されると結論づけた。本研究の考察により反映した学習者作文の特徴や問題に対し、中国語教育における作文指導のあり方について、次の 3 点の提言を行う。

一つ目は、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べきであると思われる。

二つ目は、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。

三つ目は、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。

8.3 本章のまとめ

本章では、第 6 章と第 7 章の考察結果を踏まえ、中国語母語話者の作文による言語表現との比較で、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た。また、学習者の中国語作文における特徴と問題点について、学習者の作文教科書に使用される言語表現を調べることで原因を探った。最後に、中国語教育、特に中国語作文指導に本研究の成果がいかに適用されうるか提言を行った。

第 9 章　結び

本章では、各章の関連性をまとめた上で、本論文の意義を述べ、本研究の不足点及び今後取り組むべき課題を挙げる。

9.1 本論のまとめ

各章で述べた内容及び得られた考察結果は、以下のようにまとめられる。

第 1 章はイントロダクション（introduction）として本論の研究背景と目的を紹介した。続いて、第 2 章では、本論で用いた研究手法である多次元分析に関連する先行研究を整理した上で、本論の位置付けを示した。

第 3 章と第 4 章は、本研究で行った多次元分析の基礎となるものである。第 3 章では、変数として多次元分析に用いる言語項目（Linguistic Features）を定めた。すなわち、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した。また、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した。第 4 章では、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した。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実際に検索した場合の手順と結果例を説明した。さらに、第 4 章で紹介した作文コーパスのテキストにおける各言語項目の出現頻度を、第 3 章で開発したプログラムで集計することができた。

第 5 章～第 7 章は、本論の中心的部分である。第 5 章では、第 3 章と第 4 章で得られた相対頻度表を用い、因子分析を行った。その結果、言語変異を反映しうる 4 つの共起パターン（因子）を特定することができた。次に、第 6 章では、第 5 章で得られた 4 つの共起パターンを主な次元として捉え、各次元における言語項目の共通的伝達機能に基づき、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、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。第 7 章では、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を調べた。

第 8 章では、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、中国語教育向けの提言を行った。最後に、本章では、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言及する。

9.2 本論の意義

本論の研究意義は 4 点ある。

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、及びこれらの特徴と問題点に基づく中国語教育コンテンツに向けた提言は、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。多次元分析法は、L2 英語ライティングの言語変異に関する研究の進展が著しいが、L2 中国語ライティングの研究分野では、日本人中国語学習者の中国語叙述文における言語変異を調査した徐（2021）の予備的研究以外、管見の限りでは、多次元分析法による中国語学習者の作文における言語変異に関する体系的研究が見当たらない。これにより、本研究は、L2 中国語ライティングにおける言語変異の研究ギャップを埋めるものであると言える。

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。

9.3 今後の課題

本論では、中国語母語話者の作文を比較対象とし、多次元分析法に関わる言語変異の研究視点から、日本人中国語学習者が書いた中国語作文における言語変異の考察を行った。これらの調査で得られた知見は中国語教育に資することができるが、課題も残っている。本論の不足点及び今後の課題は大別して以下の 3 点がある。

1 点目は、次元解釈の問題である。本研究では、次元及び次元に属する言語項目は量的統計手法（因子分析）で客観的に抽出されたものであるが、次元の解釈に関して、主に質的分析の視点から、次元を構成する言語項目の共通的伝達機能に基づいてラベリングしたものであるため、ラベリングから全て恣意性を排除することはできない。Yong & Pearce（2013）が因子分析手法の制約の一つとして述べているように、次元のラベリングが問題になる可能性がある[88]。表 6.1 でも示したように、いくつかの説明変数（言語項目）は、 1 つ以上の次元に出現するかもしれないので、解釈するのが困難であり、筆者もかつて、「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」と述べたことがある（徐 2021）。これについて、今後の研究において、更なる検討が必要である。

2 点目は、作文コーパスの均衡性の問題である。本論では、学習者が書いた 5 ジャンルの中国語作文を集めたが、表 4.1 に示すように、JP2（中級）と JP3（上級）サブコーパス間、コーパスサイズ（テキスト量）に不均衡がある面は否めない。JP2 サブコーパスに収録されたテキストは、「中級中国語作文授業」を履修している大学 2 年生の日本人が書いた作文であるのに対し、JP3 サブコーパスに収録されたテキストは、「上級中国語作文授業」を履修している大学 3 年生の日本人が書いた作文である。しかし、中級と上級中国語作文授業では、中国語作文種類（ジャンル）の指導重点が同じではない。なぜなら、「中級中国語作文授業」では、日記や人物描写、出来事を書くといった叙述文の練習に重点を置いているが、「上級中国語作文授業」では、叙述文の練習に加えて、中国語を専攻とする学生が、将来的に卒業論文の中国語要旨を書くための基礎作りとして、意見文の練習に重点を置くようにしているからである。そのため、JP2 と JP3 サブコーパスに含まれる作文のジャンル、テキスト数について、バランスが取れていない状態になってしまった。上述の不足を補うために、今後は、コーパスデータを拡大し、バ

[88] “One of the limitations of this technique is that naming the factors can be problematic. Factor names may not accurately reflect the variables within the factor” (Yong & Pearce 2013).

ランスよく作文データを集めた上で、更なる考察と詳細な分析を加えることが必要になるであろう。

3 点目は、中国語作文教育向けの提案である。本研究では、母語話者の作文と比較することで、学習者の中国語作文に見られた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の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避けた。“教学有法，教无定法，贵在得法”[89]（教えるには方法があるが、明確な方法はなく、適切な方法を使うことが重要である）と中国の有名な教育者である葉聖陶（中国語：叶圣陶）が述べているように、教える、学ぶには一定の方向性や目標（例：教えるべき・学ぶべき内容）があるが、教える時には従わなければならない方法はなく、対象や具体的な事情に応じて適切な教育（或いは教授法）を施すこと（孔子の“因材施教”という観点に似ている）が不可欠となる。ゆえに具体的な教育方法の検討は本研究の範囲を超えており、今後別のアプローチによって検討を進めていきたい。この点については、今後、異なる教授法により如何なる教育効果が生み出されるのかを実証的に調べていきたい。

[89] 教学有法：教える、学ぶには方法がある（教える、学ぶには方向性や目標がある）。

教无定法：教え方には明確な方法がない（教える時に従わなければならない方法はなく、教えることの目的を達成できる限り、あらゆる道理に適った方法を用いることが求められる）。

贵在得法：適切な方法を使うことが重要である（対象によって異なる方法を使うべきである。最良の結果を得る方法が最良の方法であり、その方法を見つけることが重要である）。

参考文献

（日本語）

石川慎一郎（2008）『英語コーパスと言語教育―データとしてのテクスト』大修館書店。

石川慎一郎（2012）『ベーシックコーパス言語学』東京：ひつじ書房。

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。

石黒圭（2008）『文章は接続詞で決まる』光文社新書。

輿水優（2005）『中国語の教え方・学び方：中国語科教育法概説』冨山房インターナショナル。

小林雄一郎（2017）『R によるやさしいテキストマイニング:機械学習編』オーム社。

斉藤俊雄・朝尾幸次郎・山崎俊次・ 新井洋一・梅咲敦子・塚本聡（2003）『（訳）コーパス言語学―言語構造と用法の研究』東京：南雲堂。

齊藤遙（2007） 「日本人學習者の中國語作文は通じるか--理解度からみる作文指導への示唆」『中国文学研究』(33), 116-131.

徐勤（2021）「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」『言語文化共同研究プロジェクト 2020』, 27-41.

徐勤（2022）「日本人中国語学習者の作文における接続詞の使用実態―中国語母語話者による作文との比較―」『言語文化学会』,123-140.

竹内理・水本篤（2012）『外国語教育研究ハンドブック: 研究手法のより良い理解のために』松柏社。

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,75-89.

田畑智司（2005）「コーパスに基づく文体論研究」齊藤他 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。

日本中国語学会 編（2022）『中国語学辞典』岩波書店。

白煜（2018）「自己表現育成のための外国語教育--中国語教育における教材と指導法を中心に」『関西外国語大学博士論文』。

服部環（2010）「現代の探索的因子分析における技術的選択肢」『筑波大学心理学研究』, 39(1), 11-24.

三潴正道（2015)「韻律から見た現代中国語白話書面語 (論説体) の特徴」 初探. 麗澤大学紀要, 98, 73-78.

楊彩虹（2022）「初級中国語作文授業の実践研究」『高等教育ジャーナル: 高等教育と生涯学習』(29), 135-143.

姚希孟（1988）「日本人学生の中国語作文に見られる日本語の干渉について」『外国語教育: 理論と実践』 (14), 15-25.

（中国語）

崔诚恩 (2002).《现代汉语情态副词研究》 (Doctoral dissertation，中国社会科学院).

董娜 (2003). 模糊限制语的界定及分类. 北京第二外国语学院学报,(4),28-34.

方喻 (2013). 关于外国学生的汉语模糊限制语习得考察及教学对策 (Master's thesis, 复旦大学).

冯胜利 (2003). 书面语语法及教学的相对独立性. 语言教学与研究, 02, 53-63.

冯胜利 (2006).《汉语书面用语初编》, 北京：北京语言大学出版社.

冯胜利 (2006). 论汉语书面正式语体的特征与教学. 世界汉语教学, (4), 98-106.

冯胜利 (2014). 语体俗、正、典三分的历史见证: 风、雅、颂. 语文研究, (2), 1-10.

冯胜利・王洁・黄梅(2008). 汉语书面语体庄雅度的自动测量. 语言科学 ,7(2) . 113-126.

高宁慧(1996). 留学生的代词偏误与代词在篇章中的使用原则. 世界汉语教学, (2), 61-71.

洪水英 (2015). 模糊限制语与汉语议论文语言分析. 齐齐哈尔大学学报: 哲学社会科学版,(2),127-130.

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.

黄丽燕・王嘉樱（2020）基于 Python 和 Coh-Metrix 高考英语“一年两考”阅读文本复杂度分析.外语测试与教学, (3), 1-11.李汛 (2009). 汉语综合写作教程, 北京：北京大学出版社.

凌远征 (2000).《现代汉语缩略语》. 语文出版社.

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. 江汉学术, 38 (3) , 100-110.

刘艳春・胡凤国 (2011). “播新闻” 与 “说新闻” 语体比较研究——以中央电视台《朝闻天下》 和《 马斌读报》 为例. 语言教学与研究, (1), 97-104.

柳燕妮 (2007).《副词的分类与释义分析》 (Doctoral dissertation，西北师范大学).

卢海燕(2007). 大学英语写作中人称指代使用的量化对比研究. 河南师范大学学报: 哲学社会科学版, 34(4), 157-158.

马广惠(2002). 中美大学生英语作文语言特征的对比分析. 外语教学与研究, 34(5), 45-49.

马明艳(2017). 汉语学习者书面语作文“口语化” 倾向的语体表征. 汉语学习, (1), 81-90.

亓华・贾丹丹(2013).《汉语・纵横 写作课本 （上）》. 北京：北京语言大学出版社.

宋晓娟 (2008).《时间副词用于肯定与否定的对称与不对称现象的研究》 (Master's thesis，陕西师范大学).

王保健・王宝平(2019). 英语写作人称代词过度使用及其教学对策. 内蒙古师范大学学报(教育科学版), 32(2), 99-103.

王艺璇(2015). HSK 作文成绩与句子长度, 复杂度及语法错误的相关性. 汉语应用语言学研究, (1), 139-148.

王永娜 (2017). 汉语书面正式语体语法的泛时空化特征研究. 海外华文教育动态, (2), 109-109.

文雅丽 (2007). 《现代汉语心理动词研究》 (Doctoral dissertation, 北京语言文化大学).

吴继峰(2016). 英语母语者汉语写作中的词汇丰富性发展研究. 世界汉语教学, 30(1), 129-142.

吴继峰(2019). 韩国学生不同文体写作中的语言特征对比研究. 语言教学与研究, (5), 1-12.

徐勤(2021). 泰国学生汉语记叙文第一人称代词及其作者身份表达特点初探. 海外华文教育 (Overseas Chinese Education) , (2), 34-45.

袁晖・李熙宗(2005). 《汉语语体概论》. 北京: 商务印书馆.

张江丽 (2018). 汉语作为第二语言学习者笔语产出性词汇研究. 世界汉语教学,32(3),417-430.

赵小东・尚晓华 (2012). 从人称代词角度考察大学生作文的口语化表现. 外语学刊, (6), 119-123.

（英語）

Aguado-Jiménez, P., Pérez-Paredes, P., & Sánchez, P. (2012). Exploring the use of multidimensional analysis of learner language to promote register awareness. *System*, 40(1), 90-103.

Ahmad, S., & Mahmood, M. A. (2015). Comparing Explicit Features of Pakistani Press Reportage with British Press Reportage: A Multi-Dimensional Analysis. *NUML Journal of Critical Inquiry*, 13(2).

Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS

university students' writing. *Automatic treatment and analysis of learner corpus data*, 249-264.

Asención-Delaney, Y., & Collentine, J. (2011). A multidimensional analysis of a written L2 Spanish corpus. *Applied linguistics*, 32(3), 299-322.

Azher, Musarrat & Faiz, Rabia & Chaudhri, Ayesha & Ali, Samina & Awan, Riffat-Un-Nisa. (2019). Revealing Disciplinary Variation in Pakistani Academic Writing: A Multidimensional Analysis. *International Journal of English Linguistics*, 9(2), 258-272.

Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, 384-414.

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.

Biber, D. (1995). Dimensions of register variation: A cross-linguistic comparison. 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.

Biber, D., & Conrad, S. (2014). Variation in English: Multi-dimensional studies. Routledge.

Biber, D., & Conrad, S. (2019). *Register, genre, and style* (Second edition). Cambridge: Cambridge University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use.* Cambridge University Press.

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL quarterly*, 36(1), 9-48.

Biber D, Conrad S, Reppen R, Byrd P, Helt M, Clark V, Cortes V, Csomay E, Urzua A. (2004). *Representing Language Use in the University: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus*. Princeton: Educational Testing Service.

Biber, D., & Egbert, J. (2016). Register variation on the searchable web: A multi-dimensional analysis. *Journal of English Linguistics*, 44(2), 95-137.

Biber, D., & Finegan, E. (1989). Drift and the evolution of English style: A history of three genres. *Language*, 487-517.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar*

of spoken and written English (Vol. 2). London: Longman.

Carroll, J. B. (1960). Vectors of prose style. *Style in language*, 283-292.

Chafe, W. (1985). Linguistic differences produced by differences between speaking and writing. *Literacy, language, and learning: The nature and consequences of reading and writing*, 105, 105-123.

Christie, F. (1985). Language and schooling. *Language, schooling and society*, 21-40.

Cliff, N. (1987). *Analyzing multivariate data*. Harcourt Brace Jovanovich.

Dewaele, J. M. (2018). Why the dichotomy 'L1 versus LX user' is better than 'native versus non-native speaker'. *Applied Linguistics*, 39(2), 236–240.

Douglas, D. (1992). The multi-dimensional approach to linguistic analyses of genre variation: An overview of methodology and findings. *Computers and the Humanities*, 26(5), 331-345.

Dzurjuk, T. (2006). Sentence length as a feature of style (applied to works of German writers). *Glottometrics*, (12), 55–62.

Egbert, J. (2012). Style in nineteenth century fiction: A multi-dimensional analysis. *Scientific Study of Literature*, 2(2), 167-198.

Egbert,J. (2015). Publication type and discipline variation in published academic writing: Investigating statistical interaction in corpus data. *International Journal of Corpus Linguistics*, 20(1), 1-29.

Egbert, J. (2016). Linguistic variation in research articles: When discipline tells only part of the story. *English for Specific Purposes*, 100(44), 86-87.

Egbert, J., & Biber, D. (2018). Do all roads lead to Rome?: Modeling register variation with factor analysis and discriminant analysis. *Corpus Linguistics and Linguistic Theory*, 14(2), 233-273.

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford University Press.

Field, A. (2009). *Discovering statistics using SPSS* (3th ed.). London: SAGE.

Friginal, E., & Weigle, S. (2014). Exploring multiple profiles of L2 writing using multi-dimensional analysis. *Journal of Second Language Writing*, 26, 80-95.

Friginal, E., & Mustafa, S. S. (2017). A comparison of US-based and Iraqi English research article abstracts using corpora. *Journal of English for Academic Purposes*, 25, 45-57.

Gardner, S., Biber, D., & Nesi, H. (2015). MDA perspectives on Discipline and Level in the BAWE corpus. *Corpus Linguistics 2015*, 255(229), 126.

Gorsuch, Richard. (1983). *Factor Analysis, 2nd edition*. Hillsdale, NJ: Lawrence Erlbaum Association.

Goulart, L. (2021). Register variation in L1 and L2 student writing: A multidimensional analysis. *Register Studies*, 3(1), 115-143.

Gray, B. (2013). More than discipline: Uncovering multi-dimensional patterns of variation in academic research articles. *Corpora*, 8(2), 153-181.

Grieve, J., Biber, D., Friginal, E., & Nekrasova, T. (2010). Variation among blogs: A multi-dimensional analysis. *In Genres on the Web* (pp. 303-322). Springer, Dordrecht.

Halliday, M.A.K., & Hasan, R. (1989). *Language, context, and text: Aspects of language in a social –semiotic perspective* (2nd ed.). Oxford: Oxford University Press.

Hardy, J. A., & Friginal, E. (2012). Filipino and American online communication and linguistic variation. *World Englishes*, 31(2), 143-161.

Hardy, J. A., & Friginal, E. (2016). Genre variation in student writing: A multi-dimensional analysis. *Journal of English for Academic Purposes*, 22, 119-131.

Hardy, J. A., & Römer, U. (2013). Revealing disciplinary variation in student writing: A multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 8(2), 183-207.

Heino, P. (2010). Adverbial connectors in advanced EFL learners' and native speakers' student writing. Student Thesis, University of Stockholm, Stockholm.

Hou, R., Huang, C. R., & Liu, H. (2019). A study on Chinese register characteristics based on regression analysis and text clustering. *Corpus Linguistics and Linguistic Theory*, 15(1), 1-37.

Huang, Y., & Ren, W. (2020). A novel multidimensional analysis of writing styles of editorials from China Daily and The New York Times. *Lingua*, 235, 1-16.

Jarvis, S. (2002) Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing,* 19(1):57-84.

Jiang, W. (2013). Measurements of development in L2 written production: The case of L2 Chinese. *Applied Linguistics*, 34(1), 1-24.

Jin, B. (2018). A multidimensional analysis of research article discussion sections in the field of chemical engineering. *IEEE Transactions on Professional Communication*, 61(3), 242-256.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36.

Kanoksilapatham, B. (2007). Rhetorical moves in biochemistry research articles. *Discourse on the move: Using corpus analysis to describe discourse structure*, 73-119.

Kim, J. E., & Nam, H. (2019). How do textual features of L2 argumentative essays differ across proficiency levels? A multidimensional cross-sectional study. *Reading and Writing*, 32(9), 2251-2279.

Larsen-Walker, M. (2017). Can Data Driven Learning address L2 writers' habitual errors with English linking adverbials? *System*, 69, 26-37.

Lee, D. Y. (2001). Genres, registers, text types, domains and styles: Clarifying the concepts and nevigating a path through the BNC jungle. *Language Learning & Technology*, 5(3), 37-72.

Lingard, H. C., & Rowlinson, S. (2006). Sample size in factor analysis: why size matters. *Retrieved in data*, 3, 01-12.

Liu, X., & Furneaux, C. (2014). A multidimensional comparison of discourse organization in English and Chinese university students' argumentative writing. *International Journal of Applied Linguistics*, 24(1), 74-96.

Mannion, D., & Dixon, P. (2004). Sentence-length and authorship attribution: the case of Oliver Goldsmith. *Literary and linguistic computing*, 19(4), 497-508.

Martin, J.R. (1992). *English text*. Philadelphia: John Benjamins.

McCrostie, J. (2008). Writer visibility in EFL learner academic writing: A corpus-based study. *ICAME Journal*, 32(1), 97-114

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

McEnery, A., & Xiao, Z. (2004). The Lancaster Corpus of Mandarin Chinese: A corpus for monolingual and contrastive language study. *Religion*, 17, 3-4.

Morton, A. Q. (1965). The authorship of Greek prose. Journal of the Royal Statistical Society. *Series A (General)*, 128(2), 169-233.

Myles, F., & Mitchell, R. (2004). Using information technology to support empirical SLA research. *Journal of Applied Linguistics*, 1(2).

Narita, M., Sato, C., & Sugiura, M. (2004). Connector Usage in the English Essay Writing of Japanese EFL Learners. *Proceedings of the 4th International Conference on Language Resources and Evaluation*, LREC 2004, 1, 1171–1174.

Natsukari, S. (2012). Use of I in essays by Japanese EFL learners. *JALT Journal*, 34(1), 61-78.

Omidian, T., Siyanova-Chanturia, A., & Biber, D. (2021). A new multidimensional model of writing for research publication: An analysis of disciplinarity, intra-textual variation, and L1 versus LX expert writing. *Journal of English for Academic Purposes*, 53, 101020.

Oi, K. (1999). Comparison of argumentative styles: Japanese college students vs. American college students—An analysis using the Toulmin model. *JACET Bulletin*,1999, (30), 85-102.

Pan, F. (2018). A multidimensional analysis of L1–L2 differences across three advanced levels. *Southern African Linguistics and Applied Language Studies*, 36(2), 117-131.

Parry, S. (1998). Disciplinary discourse in doctoral theses. *Higher Education*, 36(3), 273-299.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, Design and Analysis: An Integrated Approach*. Lawrence Erlbaum Associates. Inc, New Jersey.

Petch-Tyson, S. (2014). Writer/reader visibility in EFL written discourse. *Learner English on computer*, 107-118.

Pituch, K. A., & Stevens, J. P. (2016). Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS (6th ed.). Routledge: Taylor & Francis Group.

Plonsky, L., & Gonulal, T. (2015). Methodological synthesis in quantitative L2 research: A review of reviews and a case study of exploratory factor analysis. *Language Learning*, 65(S1), 9-36.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding statistics: Statistical issues in psychology, education, and the social sciences*, 2(1), 13-43.

Sardinha, T. B. (2018). Dimensions of variation across Internet registers. *International Journal of Corpus Linguistics*, 23(2), 125-157.

Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and education*, 12(4), 431-459.

Thompson, P., Hunston, S., Murakami, A., & Vajn, D. (2017). Multi-dimensional analysis, text constellations, and interdisciplinary discourse. *International Journal of Corpus Linguistics*, 22(2), 153-186.

Van Hout, R., & Vermeer, A. (2007) Comparing measures of lexical richness. In: H. Daller, J. Milton & J. Treffers-Daller (eds.), *Modelling and assessing vocabulary knowledge*, 121–138, Amsterdam: Benjamins.

Verspoor, M., Schmid, M. S., & Xu, X. (2012) A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, 21(3): 239–263.

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, 44(3), 219-246.

Weigle, S. C., & Friginal, E. (2015). Linguistic dimensions of impromptu test essays compared with successful student disciplinary writing: Effects of language background, topic, and L2 proficiency. *Journal of English for Academic Purposes*, 18, 25-39.

Wu, Y. (2005). A comparative study of news and novel style. *Chinese Monthly*, (5), 66–67.

Xie, S. (2020). Multidimensional analysis of Master thesis abstracts: a diachronic perspective. *Scientometrics*, 123(2), 861-881.

Yanchun, L., Fengguo, H., Yi, Z., & Lea, Q. (2018). Comparative Study of a Multi-Dimension/Multi-Feature Approach Between Chinese Debate and Speech. *International Journal of Language and Linguistics*, 6(5), 163-172.

Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing,* 28, 53-67.

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.

参考にしたリンク：

[1] 中国語母語話者の作文ウェブサイト：
http://www.zuowen.com/gaozhong/〔最終アクセス日：2022 年 5 月 20 日〕

[2] Streamlit の API 機能:
https://docs.streamlit.io/library/api-reference〔最終アクセス日：2022 年 10 月 25 日〕

[3] factor_analyzer package：
https://factor-analyzer.readthedocs.io/en/latest/factor_analyzer.html
〔最終アクセス日：2022 年 6 月 30 日〕

[4]《新汉语水平考试(HSK)词汇(2012 年修订版)》：
https://www.chinesetest.cn/godownload.do〔最終アクセス日：2022 年 5 月 20 日〕

[5]“现代汉语语料库分词类词频表”（現代中国語コーパスの品詞頻度表）：
http://corpus.zhonghuayuwen.org/Resources.aspx〔最終アクセス日：2022 年 3 月 19 日）

JP2 _textbook（中級中国語作文授業の教科書）[90]から抽出した作文例文の題目：

《这就是我》、《我是谁》、《和她交朋友吧》、《给编导老师的信》、《给父母的信》、《给新朋友的信》、《给老朋友的信》、《中国留学日记》、《留学日记》、《我学汉语的一件趣事》、《包饺子》、《在中国学武术》、《难忘的北师大运动会》、《我的哥哥》、《我的老师》、《难忘的同屋》、《差不多先生传》

JP3 _textbook（上級中国語作文授業の教科書）[91]から抽出した作文例文の題目：

《善良的妈妈》、《我的同学--宋福》、《中国人的温暖》、《第一次去海边钓鱼》、《走自己的路》、《不要盲目模仿》、《保护大自然》、《保护我们的“母亲”》、《上大学还是开公司》、《说说“辍学创业”》、《网络的利与弊》、《一次性餐具的利与弊》

[90] 亓华・贾丹丹(2013).《汉语・纵横 写作课本 （上）》. 北京：北京语言大学出版社.

[91] 李汛(2009). 汉语综合写作教程, 北京：北京大学出版社.

付録 I　NLPIR のタグセット

注：下表は NLPIR の《计算所汉语词性标记集》により引用したものである。

1. 名词(1 个一类，7 个二类，5 个三类) 名词分为以下子类: n 名词 nr 人名 nr1 汉语姓氏 nr2 汉语名字 nrj 日语人名 nrf 音译人名 ns 地名 nsf 音译地名 nt 机构团体名 nz 其它专名 nl 名词性惯用语 ng 名词性语素
2. 时间词(1 个一类，1 个二类) t 时间词 tg 时间词性语素
3. 处所词(1 个一类) s 处所词
4. 方位词(1 个一类) f 方位词
5. 动词(1 个一类，9 个二类) v 动词 vd 副动词 vn 名动词 vshi 动词“是” vyou 动词“有” vf 趋向动词 vx 形式动词 vi 不及物动词（内动词） vl 动词性惯用语 vg 动词性语素

6. 形容词(1 个一类，4 个二类) a 形容词 ad 副形词 an 名形词 ag 形容词性语素 al 形容词性惯用语
7. 区别词(1 个一类，2 个二类) b 区别词 bl 区别词性惯用语
8. 状态词(1 个一类) z 状态词
9. 代词(1 个一类，4 个二类，6 个三类) r 代词 rr 人称代词 rz 指示代词 rzt 时间指示代词 rzs 处所指示代词 rzv 谓词性指示代词 ry 疑问代词 ryt 时间疑问代词 rys 处所疑问代词 ryv 谓词性疑问代词 rg 代词性语素
10. 数词(1 个一类，1 个二类) m 数词 mq 数量词
11. 量词(1 个一类，2 个二类) q 量词 qv 动量词 qt 时量词
12. 副词(1 个一类) d 副词
13. 介词(1 个一类，2 个二类) p 介词 pba 介词“把”

pbei 介词“被”
14. **连词(1 个一类，1 个二类)** c 连词 cc 并列连词
15. **助词(1 个一类，15 个二类)** u 助词 uzhe 着 ule 了 喽 uguo 过 ude1 的 底 ude2 地 ude3 得 usuo 所 udeng 等 等等 云云 uyy 一样 一般 似的 般 udh 的话 uls 来讲 来说 而言 说来 uzhi 之 ulian 连 （“连小学生都会”）
16. **叹词(1 个一类)** e 叹词
17. **语气词(1 个一类)** y 语气词
18. **拟声词(1 个一类)** o 拟声词
19. **前缀(1 个一类)** h 前缀
20. **后缀(1 个一类)** k 后缀
21. **字符串(1 个一类，2 个二类)** x 字符串 xe Email 字符串 xs 微博会话分隔符 xm 表情符号 xu 网址 URL

22. 标点符号(1 个一类，16 个二类)

w 标点符号

wkz 左括号，全角：（ 〔 ［ ｛ 《 【 〖 〈 半角：([{ <

wky 右括号，全角：） 〕 ］ ｝ 》 】 〗 〉 半角：)] { >

wyz 左引号，全角：“ ‘ 『

wyy 右引号，全角：” ’ 』

wj 句号，全角：。

ww 问号，全角：？ 半角：?

wt 叹号，全角：！ 半角：!

wd 逗号，全角：， 半角：,

wf 分号，全角：； 半角： ;

wn 顿号，全角：、

wm 冒号，全角：： 半角： :

ws 省略号，全角：…… …

wp 破折号，全角：—— ——— ———— 半角：--- ----

wb 百分号千分号，全角：％ ‰ 半角：%

wh 单位符号，全角：￥ ＄ ￡ ° ℃ 半角：$

付録 II　言語項目

	ID	言語項目名	言語項目名（中国語／英語）	例 (Example)
名詞類	1	高頻度名詞	高频名词 (Noun: most commonly used)	人、社会、问题
	2	中頻度名詞	中频名词 (Noun: moderately commonly used)	厨房、奖金、书籍
	3	低頻度名詞	低频名词 (Noun: rarely used)	决赛、舞会、泥浆
	4	抽象名詞	抽象名词 (Abstract noun)	奥秘、魅力、底线
	5	具象名詞	具象名词 (Concrete noun)	婴孩、酸奶、取景框
	6	心理名詞	心理名词 (Mental noun)	悲欢、廉耻、愁绪
	7	指人名詞	指人名词 (Personal noun)	书法家、工程师、保安
	8	集合名詞	集体名词 (Collective noun)	北大、财政局、编委会
	9	普通名詞の複数形	普通名词复数型 (Common noun + plural/们)	朋友<u>们</u>、老师<u>们</u>、 战友<u>们</u>
	10	名詞化機能詞	名词化功能词 (Gerundive functional word)	研究、储备、方便
動詞類	11	高頻度動詞	高频动词 (Verb: most commonly used)	说、到、看
	12	中頻度動詞	中频动词 (Verb: moderately commonly used)	修理、协助、公布
	13	低頻度動詞	低频动词 (Verb: rarely used)	交易、领悟、收费
	14	動作動詞	动作行为动词 (Action verb)	唱歌、拔掉、拆迁
	15	心理動詞	心理动词 (Mental verb)	宠爱、忏悔、揣摩
	16	肯定を表す動詞	肯定动词 (Indicative verb)	认可、查明、证实
	17	伝達動詞	交际动词 (Communicative verb)	安慰、告诫、命令
	18	推測を表す動詞	推测性动词 (Speculative verb)	觉得、猜测、推测
	19	方向動詞	趋向动词 (Directional verb)	起来、出来、上去
	20	副動詞	副动词 (Adverbial verb)	<u>持续</u>扩大、 <u>带头</u>违反纪律
	21	軽動詞	形式动词 (light verb)	进行、给予、加以
	22	系動詞‘是’	系动词‘是’(Copular verb ‘be’/是)	尤其<u>是</u>、无疑<u>是</u>
形容詞類	23	高頻度形容詞	高频形容词 (Adjective: most commonly used)	大、新、多
	24	中頻度形容詞	中频形容词 (Adjective: moderately commonly used)	舒服、得意、富裕
	25	低頻度形容詞	低频形容词 (Adjective: rarely used)	流畅、友好、疲乏
	26	性質形容詞	性质形容词 (Qualitative adjective)	好、坏、聪明、优秀
	27	状態形容詞	状态词 (Stative adjective)	雪白、孤零零
	28	区別詞	区别词 (Distinguishing word)	<u>副</u>校长、<u>野生</u>动物
副詞類	29	高頻度副詞	高频副词 (Adverb: most commonly used)	也、不、都
	30	中頻度副詞	中频副词 (Adverb: moderately commonly used)	实在、不必、至少
	31	低頻度副詞	低频副词 (Adverb: rarely used)	一概、最好、否则
	32	必然性を表す副詞	必然性副词 (Adverb: Obligatory)	必须、毋庸置疑

	33	可能性を表す副詞	可能性副词 (Adverb: possibility)	也许、大概、可能
	34	態度を表す副詞	态度性副词 (Adverb: attitude)	难道、居然、多亏
	35	否定を表す副詞	否定性副词 (Adverb: negative)	不、没有、未曾
	36	時間を表す副詞	时间副词 (Adverb of time)	方才、偶尔、马上
	37	程度を表す副詞	程度副词 (Adverb of degree)	极度、极其、颇为
代名詞類	38	一人称単数代名詞“我”	第一人称单数代词“我”(First person pronoun: I)	我
	39	一人称複数代名詞“我们”	第一人称复数代词“我们”(First person pronoun: we)	我们
	40	他の一人称代名詞	其他第一人称代词 (Less frequently used first person pronoun)	咱、俺、敝人、在下
	41	二人称代名詞	第二人称代词 (Second person pronoun)	你、您、你们
	42	三人称代名詞	第三人称代词 (Third person pronoun)	她、他、她们、他们
	43	代名詞“它(们)”	无生性代词“它(们)”(Inanimate pronoun: it)	它、它们
	44	指示代名詞	指示代词 (Demonstrative pronoun)	这儿、那里、这些
	45	不定代名詞	不定代词 (Indefinite pronoun)	某人、某、任何
	46	疑問代名詞	疑问代词 (interrogative pronoun)	什么、谁、怎样
他の内容語類	47	数詞	数词 (Numeral)	一、20、千、一些、一个
	48	量詞	量词 (Quantifier)	两<u>点</u>、十<u>元</u>、六<u>只</u>
	49	方位場所詞	方位处所词 (Locative word)	上/f、窗外/s、心中/s
	50	感嘆詞	叹词 (Exclamation)	唉/e、嗨/e、哎呀/e
	51	擬音語	拟声词 (Onomatopoeia)	嘭/o、叽叽喳喳/o
	52	構造助詞“的”	结构助词“的”(Possessive affix: de)	我<u>的</u>书（私の本）、幸福的生活（幸せな生活）
	53	構造助詞“地”	结构助词“地”(Adverbializer: di)	好好(<u>地</u>)工作（ちゃんと仕事をする）、慢慢(<u>地</u>)说（ゆっくり話す）
	54	構造助詞“得”	结构助词“得”(Resultative complementizer: de)	做<u>得</u>到（やり遂げる）、起得很早（起きるのがとても早い）、听得懂（聞いて理解できる）
	55	列挙助詞“等/等等”	列举助词“等/等等”(Enumeration auxiliary words)	等、等等（…など，…などなど）
	56	時態助詞“着”	时态助词“着”(Durative aspect: zhe)	着（継続：～している）

	57	時態助詞“了”	时态助词“了”(Past aspect: le)	了（実現・完了：～した）
	58	時態助詞“过”	时态助词“过”(Experiential aspect: guo)	过（経験：～したことがある）
	59	比況助詞	比况助词 (Comparative auxiliary words)	一样、似的（…のようである）
機能語類	60	介詞	介词 (Preposition)	在/p、跟/p、对/p、向/p
	61	接続詞	连词 (Conjunction)	因为/c、不但/c、而且/c、和/cc
	62	語気詞	语气词 (Modal particle)	啊/y、呢/y、焉/y、呀/y
他の語彙類	63	口語詞	口语词 (Spoken word)	兔崽子、够呛、没辙
	64	挿入語	插入语 (Parenthesis)	我觉得、你瞧瞧
	65	省略語	缩略语 (Abbreviation)	冬奥会、博导、家教
	66	ぼかし言葉	模糊限制语 (Adverbial hedge)	至少、约摸、很、大概
	67	韻律単音節語	嵌偶单音词 (Monosyllabic word used in disyllabic template)	**极**佳、**围**敌、**返**京、雪**景**
	68	韻律双音節語	合偶双音词 (Disyllabic word used in disyllabic template)	**进行**讨论、**从事**教学、**建筑**桥梁
	69	古語詞	古语词 (Classical word)	称之、颇有、发轫于
	70	低難度語彙	低难度词汇 (Low-level vocabulary)	大、人、茶、老师、工作
	71	中難度語彙	中难度词汇 (Medium-level vocabulary)	超市、打扫、锻炼、聊天
	72	高難度語彙	高难度所词 (High-level vocabulary)	宝贵、采访、潮湿、诚恳
	73	非 HSK 語彙	非 HSK 词汇 (Non-HSK Vocabulary)	吞噬、锐利、璀璨、复苏
構文	74	“把”構文	“把”字句 (Preposition ba/把 sentence)	把/pba
	75	受身文	被动句 (Passive construal)	被/pbei
	76	受身文：動作主なし	无施事者被动句 (Passives: agentless)	小王**被骂**了
	77	過去を表す構文	过去时貌词 (Past temporal adverb)	已经/曾经/已/曾/刚/刚刚/方才+v、 v+过/了
	78	動作の進行を表す構文	进行式句式 (Progressive construal)	正看/looking、玩耍中/playing
	79	並列複文	并列复句 (Parallel compound sentence)	一边 A……一边 B（A をしながら B をする）；不是 A……而是 B……（A ではなく B である）

	80	連続複文	顺承复句 (Successive compound sentence)	首先 A……然后 B（まず A、それから B）；一 A……就 B……（A をすれば、すぐに B）
	81	解説複文	解说复句 (Explanatory compound sentence)	即 A……；就是说 A……（つまり A）
	82	選択複文	选择复句 (Selective compound sentence)	宁可 A……也不 B（たとえ A、むしろ B）；或者 A……或者 B……（A か B か）
	83	累加複文	递进复句 (Progressive compound sentence)	不但 A……而且 B（A だけでなく B も）；尚且 A……何况 B……（…でさえ…ではなおさらである）
	84	条件複文	条件复句 (Conditional compound sentence)	只要 A……就 B（A をしさえすれば B である）；不管 A……都 B……（A にかかわらず B だ）
	85	仮定複文	假设复句 (Hypothetical compound sentence)	如果 A……就 B（もし A ならば B）；即使 A……也 B……（たとえ…であっても）
	86	因果複文	因果复句 (Causal compound sentence)	因为 A……所以 B（A なので B だ）；之所以 A……是因为 B……（…であるわけは…に原因する）
	87	目的複文	目的复句 (Purpose compound sentence)	以便 A……（A のために B）；为的是 A……（…ためである）
	88	逆接複文	转折复句 (Turning compound sentence)	虽然 A……但是 B（A だけれども B だ）；尽管 A……可是 B……（…ではあるけれども）
基本項目	89	語彙多様性	词汇多样性 (Lexical diversity)	Types / √Tokens
	90	語彙密度	词汇密度 (Lexical density)	content words / Tokens
	91	単語の平均長さ	平均词长 (Average word length)	total characters / Tokens
	92	平均文長	平均句长 (Average sentence length)	total characters / total sentences

付録 III　HSK 単語リスト（計 5000 語）[92]

HSK 1 級（150 語）

No.	単語	No.	単語	No.	単語	No.	単語	No.	単語
1	爱	31	高兴	61	买	91	少	121	小
2	八	32	个	62	猫	92	谁	122	小姐
3	爸爸	33	工作	63	没关系	93	什么	123	些
4	杯子	34	狗	64	没有	94	十	124	写
5	北京	35	汉语	65	米饭	95	时候	125	谢谢
6	本	36	好	66	名字	96	是	126	星期
7	不	37	号	67	明天	97	书	127	学生
8	不客气	38	喝	68	哪	98	水	128	学习
9	菜	39	和	69	哪儿	99	水果	129	学校
10	茶	40	很	70	那	100	睡觉	130	一
11	吃	41	后面	71	呢	101	说	131	一点儿
12	出租车	42	回	72	能	102	四	132	衣服
13	打电话	43	会	73	你	103	岁	133	医生
14	大	44	几	74	年	104	他	134	医院
15	的	45	家	75	女儿	105	她	135	椅子
16	点	46	叫	76	朋友	106	太	136	有
17	电脑	47	今天	77	漂亮	107	天气	137	月
18	电视	48	九	78	苹果	108	听	138	再见
19	电影	49	开	79	七	109	同学	139	在
20	东西	50	看	80	前面	110	喂（叹词）	140	怎么
21	都	51	看见	81	钱	111	我	141	怎么样
22	读	52	块	82	请	112	我们	142	这
23	对不起	53	来	83	去	113	五	143	中国
24	多	54	老师	84	热	114	喜欢	144	中午
25	多少	55	了	85	人	115	下	145	住
26	儿子	56	冷	86	认识	116	下午	146	桌子
27	二	57	里	87	三	117	下雨	147	字
28	饭店	58	六	88	商店	118	先生	148	昨天
29	飞机	59	妈妈	89	上	119	现在	149	坐
30	分钟	60	吗	90	上午	120	想	150	做

[92] 付録 III は HSK の語彙リスト（中国語名：《新汉语水平考试(HSK)词汇(2012 年修订版)》）に基づいて整理したものである。
URL: https://www.chinesetest.cn/godownload.do〔最終アクセス日：2022 年 5 月 20 日〕

HSK 2 級（150 語）

No.	単語	No.	単語	No.	単語	No.	単語	No.	単語
151	吧	181	哥哥	211	累	241	身体	271	雪
152	白	182	给	212	离	242	生病	272	颜色
153	百	183	公共汽车	213	两	243	生日	273	眼睛
154	帮助	184	公司	214	零	244	时间	274	羊肉
155	报纸	185	贵	215	路	245	事情	275	药
156	比	186	过(助词)	216	旅游	246	手表	276	要
157	别	187	还(副词)	217	卖	247	手机	277	也
158	宾馆	188	孩子	218	慢	248	说话	278	一起
159	长(形容词)	189	好吃	219	忙	249	送	279	一下
160	唱歌	190	黑	220	每	250	虽然……但是……	280	已经
161	出	191	红	221	妹妹	251	它	281	意思
162	穿	192	火车站	222	门	252	踢足球	282	因为……所以……
163	次	193	机场	223	面条	253	题	283	阴
164	从	194	鸡蛋	224	男	254	跳舞	284	游泳
165	错	195	件	225	您	255	外	285	右边
166	打篮球	196	教室	226	牛奶	256	完	286	鱼
167	大家	197	姐姐	227	女	257	玩	287	远
168	到	198	介绍	228	旁边	258	晚上	288	运动
169	得(助词)	199	进	229	跑步	259	往	289	再
170	等(动词)	200	近	230	便宜	260	为什么	290	早上
171	弟弟	201	就	231	票	261	问	291	丈夫
172	第一	202	觉得	232	妻子	262	问题	292	找
173	懂	203	咖啡	233	起床	263	西瓜	293	着
174	对(形容词)	204	开始	234	千	264	希望	294	真
175	对(介词)	205	考试	235	铅笔	265	洗	295	正在
176	房间	206	可能	236	晴	266	小时	296	知道
177	非常	207	可以	237	去年	267	笑	297	准备
178	服务员	208	课	238	让	268	新	298	走
179	高	209	快	239	日	269	姓	299	最
180	告诉	210	快乐	240	上班	270	休息	300	左边

HSK 3 級（300 語）

No.	単語	No.	単語	No.	単語	No.	単語	No.	単語
301	阿姨	361	饿	421	接	481	其实	541	向
302	啊	362	耳朵	422	街道	482	其他	542	像
303	矮	363	发	423	节目	483	奇怪	543	小心
304	爱好	364	发烧	424	节日	484	骑	544	校长
305	安静	365	发现	425	结婚	485	起飞	545	新闻
306	把	366	方便	426	结束	486	起来	546	新鲜
307	班	367	放	427	解决	487	清楚	547	信用卡
308	搬	368	放心	428	借	488	请假	548	行李箱
309	办法	369	分	429	经常	489	秋	549	熊猫
310	办公室	370	附近	430	经过	490	裙子	550	需要
311	半	371	复习	431	经理	491	然后	551	选择
312	帮忙	372	干净	432	久	492	热情	552	要求
313	包	373	感冒	433	旧	493	认为	553	爷爷
314	饱	374	感兴趣	434	句子	494	认真	554	一般
315	北方	375	刚才	435	决定	495	容易	555	一边
316	被	376	个子	436	可爱	496	如果	556	一定
317	鼻子	377	根据	437	渴	497	伞	557	一共
318	比较	378	跟	438	刻	498	上网	558	一会儿
319	比赛	379	更	439	客人	499	生气	559	一样
320	笔记本	380	公斤	440	空调	500	声音	560	一直
321	必须	381	公园	441	口	501	世界	561	以前
322	变化	382	故事	442	哭	502	试	562	音乐
323	别人	383	刮风	443	裤子	503	瘦	563	银行
324	冰箱	384	关	444	筷子	504	叔叔	564	饮料
325	不但……而且……	385	关系	445	蓝	505	舒服	565	应该
326	菜单	386	关心	446	老	506	树	566	影响
327	参加	387	关于	447	离开	507	数学	567	用
328	草	388	国家	448	礼物	508	刷牙	568	游戏
329	层	389	过(动词)	449	历史	509	双	569	有名
330	差	390	过去	450	脸	510	水平	570	又
331	超市	391	还是	451	练习	511	司机	571	遇到
332	衬衫	392	害怕	452	辆	512	太阳	572	元
333	成绩	393	黑板	453	聊天	513	特别	573	愿意
334	城市	394	后来	454	了解	514	疼	574	月亮
335	迟到	395	护照	455	邻居	515	提高	575	越
336	除了	396	花(名词)	456	留学	516	体育	576	站
337	船	397	花(动词)	457	楼	517	甜	577	张

338	春	398	画	458	绿	518	条	578	长(动词)
339	词典	399	坏	459	马	519	同事	579	着急
340	聪明	400	欢迎	460	马上	520	同意	580	照顾
341	打扫	401	还(动词)	461	满意	521	头发	581	照片
342	打算	402	环境	462	帽子	522	突然	582	照相机
343	带	403	换	463	米	523	图书馆	583	只(量词)
344	担心	404	黄河	464	面包	524	腿	584	只(副词)
345	蛋糕	405	回答	465	明白	525	完成	585	只有……才……
346	当然	406	会议	466	拿	526	碗	586	中间
347	地(助词)	407	或者	467	奶奶	527	万	587	中文
348	灯	408	几乎	468	南	528	忘记	588	终于
349	地方	409	机会	469	难	529	为	589	种(量词)
350	地铁	410	极	470	难过	530	为了	590	重要
351	地图	411	记得	471	年级	531	位	591	周末
352	电梯	412	季节	472	年轻	532	文化	592	主要
353	电子邮件	413	检查	473	鸟	533	西	593	注意
354	东	414	简单	474	努力	534	习惯	594	自己
355	冬	415	见面	475	爬山	535	洗手间	595	自行车
356	动物	416	健康	476	盘子	536	洗澡	596	总是
357	短	417	讲	477	胖	537	夏	597	嘴
358	段	418	教	478	皮鞋	538	先	598	最后
359	锻炼	419	角	479	啤酒	539	相信	599	最近
360	多么	420	脚	480	瓶子	540	香蕉	600	作业

HSK 4 級（600 語）

No.	単語	No.	単語	No.	単語	No.	単語	No.	単語
601	爱情	721	符合	841	举办	961	仍然	1081	信心
602	安排	722	父亲	842	举行	962	日记	1082	兴奋
603	安全	723	付款	843	拒绝	963	入口	1083	行
604	按时	724	负责	844	距离	964	散步	1084	醒
605	按照	725	复印	845	聚会	965	森林	1085	幸福
606	百分之	726	复杂	846	开玩笑	966	沙发	1086	性别
607	棒	727	富	847	开心	967	伤心	1087	性格
608	包子	728	改变	848	看法	968	商量	1088	修理
609	保护	729	干杯	849	考虑	969	稍微	1089	许多
610	保证	730	赶	850	烤鸭	970	勺子	1090	学期
611	报名	731	敢	851	科学	971	社会	1091	压力
612	抱	732	感动	852	棵	972	申请	1092	呀
613	抱歉	733	感觉	853	咳嗽	973	深	1093	牙膏
614	倍	734	感情	854	可怜	974	甚至	1094	亚洲
615	本来	735	感谢	855	可是	975	生活	1095	严格
616	笨	736	干	856	可惜	976	生命	1096	严重
617	比如	737	刚	857	客厅	977	生意	1097	研究
618	毕业	738	高速公路	858	肯定	978	省	1098	盐
619	遍	739	胳膊	859	空	979	剩	1099	眼镜
620	标准	740	各	860	空气	980	失败	1100	演出
621	表格	741	工资	861	恐怕	981	失望	1101	演员
622	表示	742	公里	862	苦	982	师傅	1102	阳光
623	表演	743	功夫	863	矿泉水	983	十分	1103	养成
624	表扬	744	共同	864	困	984	实际	1104	样子
625	饼干	745	购物	865	困难	985	实在	1105	邀请
626	并且	746	够	866	垃圾桶	986	使	1106	要是
627	博士	747	估计	867	拉	987	使用	1107	钥匙
628	不得不	748	鼓励	868	辣	988	世纪	1108	也许
629	不管	749	故意	869	来不及	989	是否	1109	叶子
630	不过	750	顾客	870	来得及	990	适合	1110	页
631	不仅	751	挂	871	来自	991	适应	1111	一切
632	部分	752	关键	872	懒	992	收	1112	以
633	擦	753	观众	873	浪费	993	收入	1113	以为
634	猜	754	管理	874	浪漫	994	收拾	1114	艺术
635	材料	755	光	875	老虎	995	首都	1115	意见
636	参观	756	广播	876	冷静	996	首先	1116	因此
637	餐厅	757	广告	877	礼拜天	997	受不了	1117	引起
638	厕所	758	逛	878	礼貌	998	受到	1118	印象

639	差不多	759	规定	879	理发	999	售货员	1119	赢
640	长城	760	国籍	880	理解	1000	输	1120	应聘
641	长江	761	国际	881	理想	1001	熟悉	1121	永远
642	尝	762	果汁	882	力气	1002	数量	1122	勇敢
643	场	763	过程	883	厉害	1003	数字	1123	优点
644	超过	764	海洋	884	例如	1004	帅	1124	优秀
645	成功	765	害羞	885	俩	1005	顺便	1125	幽默
646	成为	766	寒假	886	连	1006	顺利	1126	尤其
647	诚实	767	汗	887	联系	1007	顺序	1127	由
648	乘坐	768	航班	888	凉快	1008	说明	1128	由于
649	吃惊	769	好处	889	零钱	1009	硕士	1129	邮局
650	重新	770	好像	890	另外	1010	死	1130	友好
651	抽烟	771	号码	891	留	1011	速度	1131	友谊
652	出差	772	合格	892	流利	1012	塑料袋	1132	有趣
653	出发	773	合适	893	流行	1013	酸	1133	于是
654	出生	774	盒子	894	旅行	1014	随便	1134	愉快
655	出现	775	后悔	895	律师	1015	随着	1135	与
656	厨房	776	厚	896	乱	1016	孙子	1136	羽毛球
657	传真	777	互联网	897	麻烦	1017	所有	1137	语法
658	窗户	778	互相	898	马虎	1018	台	1138	语言
659	词语	779	护士	899	满	1019	抬	1139	预习
660	从来	780	怀疑	900	毛	1020	态度	1140	原来
661	粗心	781	回忆	901	毛巾	1021	谈	1141	原谅
662	存	782	活动	902	美丽	1022	弹钢琴	1142	原因
663	错误	783	活泼	903	梦	1023	汤	1143	约会
664	答案	784	火	904	迷路	1024	糖	1144	阅读
665	打扮	785	获得	905	密码	1025	躺	1145	云
666	打扰	786	积极	906	免费	1026	趟	1146	允许
667	打印	787	积累	907	秒	1027	讨论	1147	杂志
668	打招呼	788	基础	908	民族	1028	讨厌	1148	咱们
669	打折	789	激动	909	母亲	1029	特点	1149	暂时
670	打针	790	及时	910	目的	1030	提	1150	脏
671	大概	791	即使	911	耐心	1031	提供	1151	责任
672	大使馆	792	计划	912	难道	1032	提前	1152	增加
673	大约	793	记者	913	难受	1033	提醒	1153	占线
674	大夫	794	技术	914	内	1034	填空	1154	招聘
675	戴	795	既然	915	内容	1035	条件	1155	照
676	当	796	继续	916	能力	1036	停	1156	真正
677	当时	797	寄	917	年龄	1037	挺	1157	整理
678	刀	798	加班	918	弄	1038	通过	1158	正常
679	导游	799	加油站	919	暖和	1039	通知	1159	正好

680	到处	800	家具	920	偶尔	1040	同情	1160	正确
681	到底	801	假	921	排队	1041	同时	1161	正式
682	倒	802	价格	922	排列	1042	推	1162	证明
683	道歉	803	坚持	923	判断	1043	推迟	1163	之
684	得意	804	减肥	924	陪	1044	脱	1164	支持
685	得（助动词）	805	减少	925	批评	1045	袜子	1165	知识
686	登机牌	806	建议	926	皮肤	1046	完全	1166	直接
687	等（助词）	807	将来	927	脾气	1047	网球	1167	值得
688	低	808	奖金	928	篇	1048	网站	1168	职业
689	底	809	降低	929	骗	1049	往往	1169	植物
690	地点	810	降落	930	乒乓球	1050	危险	1170	只好
691	地球	811	交	931	平时	1051	卫生间	1171	只要
692	地址	812	交流	932	破	1052	味道	1172	指
693	调查	813	交通	933	葡萄	1053	温度	1173	至少
694	掉	814	郊区	934	普遍	1054	文章	1174	质量
695	丢	815	骄傲	935	普通话	1055	污染	1175	重
696	动作	816	饺子	936	其次	1056	无	1176	重点
697	堵车	817	教授	937	其中	1057	无聊	1177	重视
698	肚子	818	教育	938	气候	1058	无论	1178	周围
699	短信	819	接受	939	千万	1059	误会	1179	主意
700	对话	820	接着	940	签证	1060	西红柿	1180	祝贺
701	对面	821	节	941	敲	1061	吸引	1181	著名
702	对于	822	节约	942	桥	1062	咸	1182	专门
703	儿童	823	结果	943	巧克力	1063	现金	1183	专业
704	而	824	解释	944	亲戚	1064	羡慕	1184	转
705	发生	825	尽管	945	轻	1065	相反	1185	赚
706	发展	826	紧张	946	轻松	1066	相同	1186	准确
707	法律	827	进行	947	情况	1067	香	1187	准时
708	翻译	828	禁止	948	穷	1068	详细	1188	仔细
709	烦恼	829	京剧	949	区别	1069	响	1189	自然
710	反对	830	经济	950	取	1070	橡皮	1190	自信
711	方法	831	经历	951	全部	1071	消息	1191	总结
712	方面	832	经验	952	缺点	1072	小吃	1192	租
713	方向	833	精彩	953	缺少	1073	小伙子	1193	最好
714	房东	834	景色	954	却	1074	小说	1194	尊重
715	放弃	835	警察	955	确实	1075	笑话	1195	左右
716	放暑假	836	竞争	956	然而	1076	效果	1196	作家
717	放松	837	竟然	957	热闹	1077	心情	1197	作用
718	份	838	镜子	958	任何	1078	辛苦	1198	作者
719	丰富	839	究竟	959	任务	1079	信封	1199	座
720	否则	840	举	960	扔	1080	信息	1200	座位

HSK 5 級（1300 語）

No.	単語	No.	単語	No.	単語	No.	単語	No.	単語
1201	哎	1461	繁荣	1721	精力	1981	人类	2241	象棋
1202	唉	1462	反而	1722	精神	1982	人民币	2242	象征
1203	爱护	1463	反复	1723	酒吧	1983	人生	2243	消费
1204	爱惜	1464	反应	1724	救	1984	人事	2244	消化
1205	爱心	1465	反映	1725	救护车	1985	人物	2245	消极
1206	安慰	1466	反正	1726	舅舅	1986	人员	2246	消失
1207	安装	1467	范围	1727	居然	1987	忍不住	2247	销售
1208	岸	1468	方	1728	桔子	1988	日常	2248	小麦
1209	暗	1469	方案	1729	巨大	1989	日程	2249	小气
1210	熬夜	1470	方式	1730	具备	1990	日历	2250	孝顺
1211	把握	1471	妨碍	1731	具体	1991	日期	2251	效率
1212	摆	1472	仿佛	1732	俱乐部	1992	日用品	2252	歇
1213	办理	1473	非	1733	据说	1993	日子	2253	斜
1214	傍晚	1474	肥皂	1734	捐	1994	如何	2254	写作
1215	包裹	1475	废话	1735	决赛	1995	如今	2255	血
1216	包含	1476	分别	1736	决心	1996	软	2256	心理
1217	包括	1477	分布	1737	角色	1997	软件	2257	心脏
1218	薄	1478	分配	1738	绝对	1998	弱	2258	欣赏
1219	宝贝	1479	分手	1739	军事	1999	洒	2259	信号
1220	宝贵	1480	分析	1740	均匀	2000	嗓子	2260	信任
1221	保持	1481	纷纷	1741	卡车	2001	色彩	2261	行动
1222	保存	1482	奋斗	1742	开发	2002	杀	2262	行人
1223	保留	1483	风格	1743	开放	2003	沙漠	2263	行为
1224	保险	1484	风景	1744	开幕式	2004	沙滩	2264	形成
1225	报到	1485	风俗	1745	开水	2005	傻	2265	形容
1226	报道	1486	风险	1746	砍	2006	晒	2266	形式
1227	报告	1487	疯狂	1747	看不起	2007	删除	2267	形势
1228	报社	1488	讽刺	1748	看望	2008	闪电	2268	形象
1229	抱怨	1489	否定	1749	靠	2009	扇子	2269	形状
1230	背	1490	否认	1750	颗	2010	善良	2270	幸亏
1231	悲观	1491	扶	1751	可见	2011	善于	2271	幸运
1232	背景	1492	服装	1752	可靠	2012	伤害	2272	性质
1233	被子	1493	幅	1753	可怕	2013	商品	2273	兄弟
1234	本科	1494	辅导	1754	克	2014	商务	2274	胸
1235	本领	1495	妇女	1755	克服	2015	商业	2275	休闲
1236	本质	1496	复制	1756	刻苦	2016	上当	2276	修改
1237	比例	1497	改革	1757	客观	2017	蛇	2277	虚心
1238	彼此	1498	改进	1758	课程	2018	舍不得	2278	叙述

1239	必然	1499	改善	1759	空间	2019	设备	2279	宣布
1240	必要	1500	改正	1760	空闲	2020	设计	2280	宣传
1241	毕竟	1501	盖	1761	控制	2021	设施	2281	学历
1242	避免	1502	概括	1762	口味	2022	射击	2282	学术
1243	编辑	1503	概念	1763	夸	2023	摄影	2283	学问
1244	鞭炮	1504	干脆	1764	夸张	2024	伸	2284	寻找
1245	便	1505	干燥	1765	会计	2025	身材	2285	询问
1246	辩论	1506	赶紧	1766	宽	2026	身份	2286	训练
1247	标点	1507	赶快	1767	昆虫	2027	深刻	2287	迅速
1248	标志	1508	感激	1768	扩大	2028	神话	2288	押金
1249	表达	1509	感受	1769	辣椒	2029	神秘	2289	牙齿
1250	表面	1510	感想	1770	拦	2030	升	2290	延长
1251	表明	1511	干活儿	1771	烂	2031	生产	2291	严肃
1252	表情	1512	钢铁	1772	朗读	2032	生动	2292	演讲
1253	表现	1513	高档	1773	劳动	2033	生长	2293	宴会
1254	冰激凌	1514	高级	1774	劳驾	2034	声调	2294	阳台
1255	病毒	1515	搞	1775	老百姓	2035	绳子	2295	痒
1256	玻璃	1516	告别	1776	老板	2036	省略	2296	样式
1257	播放	1517	格外	1777	老婆	2037	胜利	2297	腰
1258	脖子	1518	隔壁	1778	老实	2038	失眠	2298	摇
1259	博物馆	1519	个别	1779	老鼠	2039	失去	2299	咬
1260	补充	1520	个人	1780	姥姥	2040	失业	2300	要不
1261	不安	1521	个性	1781	乐观	2041	诗	2301	业务
1262	不得了	1522	各自	1782	雷	2042	狮子	2302	业余
1263	不断	1523	根	1783	类型	2043	湿润	2303	夜
1264	不见得	1524	根本	1784	冷淡	2044	石头	2304	一辈子
1265	不耐烦	1525	工厂	1785	厘米	2045	时差	2305	一旦
1266	不然	1526	工程师	1786	离婚	2046	时代	2306	一律
1267	不如	1527	工具	1787	梨	2047	时刻	2307	一再
1268	不要紧	1528	工人	1788	理论	2048	时髦	2308	一致
1269	不足	1529	工业	1789	理由	2049	时期	2309	依然
1270	布	1530	公布	1790	力量	2050	时尚	2310	移动
1271	步骤	1531	公开	1791	立即	2051	实话	2311	移民
1272	部门	1532	公平	1792	立刻	2052	实践	2312	遗憾
1273	财产	1533	公寓	1793	利润	2053	实习	2313	疑问
1274	采访	1534	公元	1794	利息	2054	实现	2314	乙
1275	采取	1535	公主	1795	利益	2055	实验	2315	以及
1276	彩虹	1536	功能	1796	利用	2056	实用	2316	以来
1277	踩	1537	恭喜	1797	连忙	2057	食物	2317	亿
1278	参考	1538	贡献	1798	连续	2058	使劲儿	2318	义务
1279	参与	1539	沟通	1799	联合	2059	始终	2319	议论
1280	惭愧	1540	构成	1800	恋爱	2060	士兵	2320	意外

1281	操场	1541	姑姑	1801	良好	2061	市场	2321	意义
1282	操心	1542	姑娘	1802	粮食	2062	似的	2322	因而
1283	册	1543	古代	1803	亮	2063	事实	2323	因素
1284	测验	1544	古典	1804	了不起	2064	事物	2324	银
1285	曾经	1545	股票	1805	列车	2065	事先	2325	印刷
1286	叉子	1546	骨头	1806	临时	2066	试卷	2326	英俊
1287	差距	1547	鼓舞	1807	灵活	2067	收获	2327	英雄
1288	插	1548	鼓掌	1808	铃	2068	收据	2328	迎接
1289	拆	1549	固定	1809	零件	2069	手工	2329	营养
1290	产品	1550	挂号	1810	零食	2070	手术	2330	营业
1291	产生	1551	乖	1811	领导	2071	手套	2331	影子
1292	长途	1552	拐弯	1812	领域	2072	手续	2332	应付
1293	常识	1553	怪不得	1813	浏览	2073	手指	2333	应用
1294	抄	1554	关闭	1814	流传	2074	首	2334	硬
1295	超级	1555	观察	1815	流泪	2075	寿命	2335	硬件
1296	朝	1556	观点	1816	龙	2076	受伤	2336	拥抱
1297	潮湿	1557	观念	1817	漏	2077	书架	2337	拥挤
1298	吵	1558	官	1818	陆地	2078	梳子	2338	勇气
1299	吵架	1559	管子	1819	陆续	2079	舒适	2339	用功
1300	炒	1560	冠军	1820	录取	2080	输入	2340	用途
1301	车库	1561	光滑	1821	录音	2081	蔬菜	2341	优惠
1302	车厢	1562	光临	1822	轮流	2082	熟练	2342	优美
1303	彻底	1563	光明	1823	论文	2083	属于	2343	优势
1304	沉默	1564	光盘	1824	逻辑	2084	鼠标	2344	悠久
1305	趁	1565	广场	1825	落后	2085	数	2345	犹豫
1306	称	1566	广大	1826	骂	2086	数据	2346	油炸
1307	称呼	1567	广泛	1827	麦克风	2087	数码	2347	游览
1308	称赞	1568	归纳	1828	馒头	2088	摔倒	2348	有利
1309	成分	1569	规矩	1829	满足	2089	甩	2349	幼儿园
1310	成果	1570	规律	1830	毛病	2090	双方	2350	娱乐
1311	成就	1571	规模	1831	矛盾	2091	税	2351	与其
1312	成立	1572	规则	1832	冒险	2092	说不定	2352	语气
1313	成人	1573	柜台	1833	贸易	2093	说服	2353	玉米
1314	成熟	1574	滚	1834	眉毛	2094	丝绸	2354	预报
1315	成语	1575	锅	1835	媒体	2095	丝毫	2355	预订
1316	成长	1576	国庆节	1836	煤炭	2096	私人	2356	预防
1317	诚恳	1577	国王	1837	美术	2097	思考	2357	元旦
1318	承担	1578	果然	1838	魅力	2098	思想	2358	员工
1319	承认	1579	果实	1839	梦想	2099	撕	2359	原料
1320	承受	1580	过分	1840	秘密	2100	似乎	2360	原则
1321	程度	1581	过敏	1841	秘书	2101	搜索	2361	圆
1322	程序	1582	过期	1842	密切	2102	宿舍	2362	愿望

1323	吃亏	1583	哈	1843	蜜蜂	2103	随身	2363	乐器
1324	池塘	1584	海关	1844	面对	2104	随时	2364	晕
1325	迟早	1585	海鲜	1845	面积	2105	随手	2365	运气
1326	持续	1586	喊	1846	面临	2106	碎	2366	运输
1327	尺子	1587	行业	1847	苗条	2107	损失	2367	运用
1328	翅膀	1588	豪华	1848	描写	2108	缩短	2368	灾害
1329	冲	1589	好客	1849	敏感	2109	所	2369	再三
1330	充电器	1590	好奇	1850	名牌	2110	锁	2370	在乎
1331	充分	1591	合法	1851	名片	2111	台阶	2371	在于
1332	充满	1592	合理	1852	名胜古迹	2112	太极拳	2372	赞成
1333	重复	1593	合同	1853	明确	2113	太太	2373	赞美
1334	宠物	1594	合影	1854	明显	2114	谈判	2374	糟糕
1335	抽屉	1595	合作	1855	明星	2115	坦率	2375	造成
1336	抽象	1596	何必	1856	命令	2116	烫	2376	则
1337	丑	1597	何况	1857	命运	2117	逃	2377	责备
1338	臭	1598	和平	1858	摸	2118	逃避	2378	摘
1339	出版	1599	核心	1859	模仿	2119	桃	2379	窄
1340	出口	1600	恨	1860	模糊	2120	淘气	2380	粘贴
1341	出色	1601	猴子	1861	模特	2121	讨价还价	2381	展开
1342	出示	1602	后背	1862	摩托车	2122	套	2382	展览
1343	出席	1603	后果	1863	陌生	2123	特色	2383	占
1344	初级	1604	呼吸	1864	某	2124	特殊	2384	战争
1345	除非	1605	忽然	1865	木头	2125	特征	2385	长辈
1346	除夕	1606	忽视	1866	目标	2126	疼爱	2386	涨
1347	处理	1607	胡说	1867	目录	2127	提倡	2387	掌握
1348	传播	1608	胡同	1868	目前	2128	提纲	2388	账户
1349	传染	1609	壶	1869	哪怕	2129	提问	2389	招待
1350	传说	1610	蝴蝶	1870	难怪	2130	题目	2390	着火
1351	传统	1611	糊涂	1871	难免	2131	体会	2391	着凉
1352	窗帘	1612	花生	1872	脑袋	2132	体贴	2392	召开
1353	闯	1613	划	1873	内部	2133	体现	2393	照常
1354	创造	1614	华裔	1874	内科	2134	体验	2394	哲学
1355	吹	1615	滑	1875	嫩	2135	天空	2395	针对
1356	词汇	1616	化学	1876	能干	2136	天真	2396	珍惜
1357	辞职	1617	话题	1877	能源	2137	调皮	2397	真实
1358	此外	1618	怀念	1878	嗯	2138	调整	2398	诊断
1359	次要	1619	怀孕	1879	年代	2139	挑战	2399	阵
1360	刺激	1620	缓解	1880	年纪	2140	通常	2400	振动
1361	匆忙	1621	幻想	1881	念	2141	统一	2401	争论
1362	从此	1622	慌张	1882	宁可	2142	痛苦	2402	争取
1363	从而	1623	黄金	1883	牛仔裤	2143	痛快	2403	征求
1364	从前	1624	灰	1884	农村	2144	偷	2404	睁

1365	从事	1625	灰尘	1885	农民	2145	投入	2405	整个
1366	粗糙	1626	灰心	1886	农业	2146	投资	2406	整齐
1367	促进	1627	挥	1887	浓	2147	透明	2407	整体
1368	促使	1628	恢复	1888	女士	2148	突出	2408	正
1369	醋	1629	汇率	1889	欧洲	2149	土地	2409	证件
1370	催	1630	婚礼	1890	偶然	2150	土豆	2410	证据
1371	存在	1631	婚姻	1891	拍	2151	吐	2411	政府
1372	措施	1632	活跃	1892	派	2152	兔子	2412	政治
1373	答应	1633	火柴	1893	盼望	2153	团	2413	挣
1374	达到	1634	伙伴	1894	培训	2154	推辞	2414	支
1375	打工	1635	或许	1895	培养	2155	推广	2415	支票
1376	打交道	1636	机器	1896	赔偿	2156	推荐	2416	执照
1377	打喷嚏	1637	肌肉	1897	佩服	2157	退	2417	直
1378	打听	1638	基本	1898	配合	2158	退步	2418	指导
1379	大方	1639	激烈	1899	盆	2159	退休	2419	指挥
1380	大厦	1640	及格	1900	碰	2160	歪	2420	至今
1381	大象	1641	极其	1901	批	2161	外公	2421	至于
1382	大型	1642	急忙	1902	批准	2162	外交	2422	志愿者
1383	呆	1643	急诊	1903	披	2163	完美	2423	制定
1384	代表	1644	集合	1904	疲劳	2164	完善	2424	制度
1385	代替	1645	集体	1905	匹	2165	完整	2425	制造
1386	贷款	1646	集中	1906	片	2166	玩具	2426	制作
1387	待遇	1647	计算	1907	片面	2167	万一	2427	治疗
1388	担任	1648	记录	1908	飘	2168	王子	2428	秩序
1389	单纯	1649	记忆	1909	拼音	2169	网络	2429	智慧
1390	单调	1650	纪录	1910	频道	2170	往返	2430	中介
1391	单独	1651	纪律	1911	平	2171	危害	2431	中心
1392	单位	1652	纪念	1912	平安	2172	威胁	2432	中旬
1393	单元	1653	系领带	1913	平常	2173	微笑	2433	种类
1394	耽误	1654	寂寞	1914	平等	2174	违反	2434	重大
1395	胆小鬼	1655	夹子	1915	平方	2175	围巾	2435	重量
1396	淡	1656	家庭	1916	平衡	2176	围绕	2436	周到
1397	当地	1657	家务	1917	平静	2177	唯一	2437	猪
1398	当心	1658	家乡	1918	平均	2178	维修	2438	竹子
1399	挡	1659	嘉宾	1919	评价	2179	伟大	2439	逐步
1400	导演	1660	甲	1920	凭	2180	尾巴	2440	逐渐
1401	导致	1661	假如	1921	迫切	2181	委屈	2441	主持
1402	岛屿	1662	假设	1922	破产	2182	未必	2442	主动
1403	倒霉	1663	假装	1923	破坏	2183	未来	2443	主观
1404	到达	1664	价值	1924	期待	2184	位于	2444	主人
1405	道德	1665	驾驶	1925	期间	2185	位置	2445	主任
1406	道理	1666	嫁	1926	其余	2186	胃	2446	主题

1407	登记	1667	坚决	1927	奇迹	2187	胃口	2447	主席
1408	等待	1668	坚强	1928	企业	2188	温暖	2448	主张
1409	等于	1669	肩膀	1929	启发	2189	温柔	2449	煮
1410	滴	1670	艰巨	1930	气氛	2190	文件	2450	注册
1411	的确	1671	艰苦	1931	汽油	2191	文具	2451	祝福
1412	敌人	1672	兼职	1932	谦虚	2192	文明	2452	抓
1413	地道	1673	捡	1933	签	2193	文学	2453	抓紧
1414	地理	1674	剪刀	1934	前途	2194	文字	2454	专家
1415	地区	1675	简历	1935	浅	2195	闻	2455	专心
1416	地毯	1676	简直	1936	欠	2196	吻	2456	转变
1417	地位	1677	建立	1937	枪	2197	稳定	2457	转告
1418	地震	1678	建设	1938	强调	2198	问候	2458	装
1419	递	1679	建筑	1939	强烈	2199	卧室	2459	装饰
1420	点心	1680	健身	1940	墙	2200	握手	2460	装修
1421	电池	1681	键盘	1941	抢	2201	屋子	2461	状况
1422	电台	1682	讲究	1942	悄悄	2202	无奈	2462	状态
1423	钓	1683	讲座	1943	瞧	2203	无数	2463	撞
1424	顶	1684	酱油	1944	巧妙	2204	无所谓	2464	追
1425	动画片	1685	交换	1945	切	2205	武术	2465	追求
1426	冻	1686	交际	1946	亲爱	2206	勿	2466	咨询
1427	洞	1687	交往	1947	亲切	2207	物理	2467	姿势
1428	豆腐	1688	浇	1948	亲自	2208	物质	2468	资格
1429	逗	1689	胶水	1949	勤奋	2209	雾	2469	资金
1430	独立	1690	角度	1950	青	2210	吸取	2470	资料
1431	独特	1691	狡猾	1951	青春	2211	吸收	2471	资源
1432	度过	1692	教材	1952	青少年	2212	戏剧	2472	紫
1433	断	1693	教练	1953	轻视	2213	系	2473	自从
1434	堆	1694	教训	1954	轻易	2214	系统	2474	自动
1435	对比	1695	阶段	1955	清淡	2215	细节	2475	自豪
1436	对待	1696	结实	1956	情景	2216	瞎	2476	自觉
1437	对方	1697	接触	1957	情绪	2217	下载	2477	自私
1438	对手	1698	接待	1958	请求	2218	吓	2478	自由
1439	对象	1699	接近	1959	庆祝	2219	夏令营	2479	自愿
1440	兑换	1700	节省	1960	球迷	2220	鲜艳	2480	字母
1441	吨	1701	结构	1961	趋势	2221	显得	2481	字幕
1442	蹲	1702	结合	1962	取消	2222	显然	2482	综合
1443	顿	1703	结论	1963	娶	2223	显示	2483	总裁
1444	多亏	1704	结账	1964	去世	2224	县	2484	总共
1445	多余	1705	戒	1965	圈	2225	现代	2485	总理
1446	朵	1706	戒指	1966	权力	2226	现实	2486	总算
1447	躲藏	1707	届	1967	权利	2227	现象	2487	总统
1448	恶劣	1708	借口	1968	全面	2228	限制	2488	总之

1449	耳环	1709	金属	1969	劝	2229	相处	2489	阻止
1450	发表	1710	尽快	1970	缺乏	2230	相当	2490	组
1451	发愁	1711	尽量	1971	确定	2231	相对	2491	组成
1452	发达	1712	紧急	1972	确认	2232	相关	2492	组合
1453	发抖	1713	谨慎	1973	群	2233	相似	2493	组织
1454	发挥	1714	尽力	1974	燃烧	2234	香肠	2494	最初
1455	发明	1715	进步	1975	绕	2235	享受	2495	醉
1456	发票	1716	进口	1976	热爱	2236	想念	2496	尊敬
1457	发言	1717	近代	1977	热烈	2237	想象	2497	遵守
1458	罚款	1718	经典	1978	热心	2238	项	2498	作品
1459	法院	1719	经商	1979	人才	2239	项链	2499	作为
1460	翻	1720	经营	1980	人口	2240	项目	2500	作文

HSK 6 級（2500 語）

No.	単語	No.	単語	No.	単語	No.	単語	No.	単語
2501	挨	3001	发扬	3501	敬业	4001	全力以赴	4501	谢绝
2502	癌症	3002	发育	3502	境界	4002	拳头	4502	心得
2503	爱不释手	3003	法人	3503	镜头	4003	犬	4503	心甘情愿
2504	爱戴	3004	番	3504	纠纷	4004	缺口	4504	心灵
2505	暧昧	3005	凡是	3505	纠正	4005	缺席	4505	心态
2506	安宁	3006	繁华	3506	酒精	4006	缺陷	4506	心疼
2507	安详	3007	繁忙	3507	救济	4007	瘸	4507	心血
2508	安置	3008	繁体字	3508	就近	4008	确保	4508	心眼儿
2509	按摩	3009	繁殖	3509	就业	4009	确立	4509	辛勤
2510	案件	3010	反驳	3510	就职	4010	确切	4510	欣慰
2511	案例	3011	反常	3511	拘留	4011	确信	4511	欣欣向荣
2512	暗示	3012	反感	3512	拘束	4012	群众	4512	新陈代谢
2513	昂贵	3013	反抗	3513	居民	4013	染	4513	新郎
2514	凹凸	3014	反馈	3514	居住	4014	嚷	4514	新娘
2515	熬	3015	反面	3515	鞠躬	4015	让步	4515	新颖
2516	奥秘	3016	反射	3516	局部	4016	饶恕	4516	薪水
2517	巴不得	3017	反思	3517	局面	4017	扰乱	4517	信赖
2518	巴结	3018	反问	3518	局势	4018	惹祸	4518	信念
2519	扒	3019	反之	3519	局限	4019	热泪盈眶	4519	信仰
2520	疤	3020	泛滥	3520	咀嚼	4020	热门	4520	信誉
2521	拔苗助长	3021	范畴	3521	沮丧	4021	人道	4521	兴隆
2522	把关	3022	贩卖	3522	举动	4022	人格	4522	兴旺

2523	把手	3023	方位	3523	举世瞩目	4023	人工	4523	腥
2524	罢工	3024	方言	3524	举足轻重	4024	人家	4524	刑事
2525	霸道	3025	方圆	3525	剧本	4025	人间	4525	行政
2526	掰	3026	方针	3526	剧烈	4026	人士	4526	形态
2527	摆脱	3027	防守	3527	据悉	4027	人为	4527	兴高采烈
2528	败坏	3028	防御	3528	聚精会神	4028	人性	4528	兴致勃勃
2529	拜访	3029	防止	3529	卷	4029	人质	4529	性感
2530	拜年	3030	防治	3530	决策	4030	仁慈	4530	性命
2531	拜托	3031	访问	3531	觉悟	4031	忍耐	4531	性能
2532	颁布	3032	纺织	3532	觉醒	4032	忍受	4532	凶恶
2533	颁发	3033	放大	3533	绝望	4033	认定	4533	凶手
2534	斑	3034	放射	3534	倔强	4034	认可	4534	汹涌
2535	版本	3035	飞禽走兽	3535	军队	4035	任命	4535	胸怀
2536	半途而废	3036	飞翔	3536	君子	4036	任性	4536	胸膛
2537	扮演	3037	飞跃	3537	卡通	4037	任意	4537	雄厚
2538	伴侣	3038	非法	3538	开采	4038	任重道远	4538	雄伟
2539	伴随	3039	肥沃	3539	开除	4039	仍旧	4539	修复
2540	绑架	3040	诽谤	3540	开阔	4040	日新月异	4540	修建
2541	榜样	3041	肺	3541	开朗	4041	日益	4541	修养
2542	磅	3042	废除	3542	开明	4042	荣幸	4542	羞耻
2543	包庇	3043	废寝忘食	3543	开辟	4043	荣誉	4543	绣
2544	包袱	3044	废墟	3544	开拓	4044	容貌	4544	嗅觉
2545	包围	3045	沸腾	3545	开展	4045	容纳	4545	须知
2546	包装	3046	分辨	3546	开支	4046	容器	4546	虚假
2547	饱和	3047	分寸	3547	刊登	4047	容忍	4547	虚荣
2548	饱经沧桑	3048	分红	3548	刊物	4048	溶解	4548	虚伪
2549	保管	3049	分解	3549	勘探	4049	融化	4549	需求
2550	保密	3050	分裂	3550	侃侃而谈	4050	融洽	4550	许可
2551	保姆	3051	分泌	3551	砍伐	4051	柔和	4551	序言
2552	保守	3052	分明	3552	看待	4052	揉	4552	畜牧
2553	保卫	3053	分歧	3553	慷慨	4053	儒家	4553	酗酒
2554	保养	3054	分散	3554	扛	4054	若干	4554	宣誓
2555	保障	3055	吩咐	3555	抗议	4055	弱点	4555	宣扬
2556	保重	3056	坟墓	3556	考察	4056	撒谎	4556	喧哗
2557	报仇	3057	粉末	3557	考古	4057	散文	4557	悬挂
2558	报酬	3058	粉色	3558	考核	4058	散布	4558	悬念
2559	报答	3059	粉碎	3559	考验	4059	散发	4559	悬殊
2560	报复	3060	分量	3560	靠拢	4060	丧失	4560	悬崖峭壁
2561	报警	3061	愤怒	3561	科目	4061	骚扰	4561	旋律
2562	报销	3062	丰满	3562	磕	4062	嫂子	4562	旋转
2563	抱负	3063	丰盛	3563	可观	4063	刹车	4563	选拔

2564	暴力	3064	丰收	3564	可口	4064	啥	4564	选举
2565	暴露	3065	风暴	3565	可恶	4065	筛选	4565	选手
2566	曝光	3066	风度	3566	可行	4066	山脉	4566	炫耀
2567	爆发	3067	风光	3567	渴望	4067	闪烁	4567	削弱
2568	爆炸	3068	风气	3568	克制	4068	擅长	4568	学说
2569	卑鄙	3069	风趣	3569	刻不容缓	4069	擅自	4569	学位
2570	悲哀	3070	风土人情	3570	客户	4070	伤脑筋	4570	雪上加霜
2571	悲惨	3071	风味	3571	课题	4071	商标	4571	血压
2572	北极	3072	封闭	3572	恳切	4072	上级	4572	熏陶
2573	贝壳	3073	封建	3573	啃	4073	上进	4573	寻觅
2574	备份	3074	封锁	3574	坑	4074	上任	4574	巡逻
2575	备忘录	3075	锋利	3575	空洞	4075	上瘾	4575	循环
2576	背叛	3076	逢	3576	空前绝后	4076	上游	4576	循序渐进
2577	背诵	3077	奉献	3577	空想	4077	尚且	4577	压迫
2578	被动	3078	否决	3578	空虚	4078	捎	4578	压岁钱
2579	被告	3079	夫妇	3579	孔	4079	梢	4579	压缩
2580	奔波	3080	夫人	3580	恐怖	4080	哨	4580	压抑
2581	奔驰	3081	敷衍	3581	恐吓	4081	奢侈	4581	压榨
2582	本能	3082	服从	3582	恐惧	4082	舌头	4582	压制
2583	本钱	3083	服气	3583	空白	4083	设立	4583	鸦雀无声
2584	本人	3084	俘虏	3584	空隙	4084	设想	4584	亚军
2585	本身	3085	符号	3585	口气	4085	设置	4585	烟花爆竹
2586	本事	3086	幅度	3586	口腔	4086	社区	4586	淹没
2587	笨拙	3087	辐射	3587	口头	4087	涉及	4587	延期
2588	崩溃	3088	福利	3588	口音	4088	摄氏度	4588	延伸
2589	甭	3089	福气	3589	扣	4089	申报	4589	延续
2590	迸发	3090	抚摸	3590	枯萎	4090	呻吟	4590	严寒
2591	蹦	3091	抚养	3591	枯燥	4091	绅士	4591	严禁
2592	逼迫	3092	俯视	3592	哭泣	4092	深奥	4592	严峻
2593	鼻涕	3093	辅助	3593	苦尽甘来	4093	深沉	4593	严厉
2594	比方	3094	腐败	3594	苦涩	4094	深情厚谊	4594	严密
2595	比喻	3095	腐烂	3595	挎	4095	神经	4595	言论
2596	比重	3096	腐蚀	3596	跨	4096	神奇	4596	岩石
2597	鄙视	3097	腐朽	3597	快活	4097	神气	4597	炎热
2598	闭塞	3098	负担	3598	宽敞	4098	神圣	4598	沿海
2599	弊病	3099	附和	3599	宽容	4099	神态	4599	掩盖
2600	弊端	3100	附件	3600	款待	4100	神仙	4600	掩护
2601	臂	3101	附属	3601	款式	4101	审查	4601	掩饰
2602	边疆	3102	复活	3602	筐	4102	审理	4602	眼光
2603	边界	3103	复兴	3603	旷课	4103	审美	4603	眼色
2604	边境	3104	副	3604	况且	4104	审判	4604	眼神

2605	边缘	3105	赋予	3605	矿产	4105	渗透	4605	演变
2606	编织	3106	富裕	3606	框架	4106	慎重	4606	演习
2607	鞭策	3107	腹泻	3607	亏待	4107	生存	4607	演绎
2608	贬低	3108	覆盖	3608	亏损	4108	生机	4608	演奏
2609	贬义	3109	改良	3609	捆绑	4109	生理	4609	厌恶
2610	扁	3110	钙	3610	扩充	4110	生疏	4610	验收
2611	变故	3111	盖章	3611	扩散	4111	生态	4611	验证
2612	变迁	3112	干旱	3612	扩张	4112	生物	4612	氧气
2613	变质	3113	干扰	3613	喇叭	4113	生肖	4613	样品
2614	便利	3114	干涉	3614	蜡烛	4114	生效	4614	谣言
2615	便条	3115	干预	3615	啦	4115	生锈	4615	摇摆
2616	便于	3116	尴尬	3616	来历	4116	生育	4616	摇滚
2617	遍布	3117	感慨	3617	来源	4117	声明	4617	遥控
2618	辨认	3118	感染	3618	栏目	4118	声势	4618	遥远
2619	辩护	3119	干劲	3619	懒惰	4119	声誉	4619	要点
2620	辩解	3120	纲领	3620	狼狈	4120	牲畜	4620	要命
2621	辩证	3121	岗位	3621	狼吞虎咽	4121	省会	4621	要素
2622	辫子	3122	港口	3622	捞	4122	胜负	4622	耀眼
2623	标本	3123	港湾	3623	牢固	4123	盛产	4623	野蛮
2624	标记	3124	杠杆	3624	牢骚	4124	盛开	4624	野心
2625	标题	3125	高超	3625	唠叨	4125	盛情	4625	液体
2626	表决	3126	高潮	3626	乐趣	4126	盛行	4626	一度
2627	表态	3127	高峰	3627	乐意	4127	尸体	4627	一帆风顺
2628	表彰	3128	高明	3628	雷达	4128	失事	4628	一贯
2629	憋	3129	高尚	3629	类似	4129	失误	4629	一举两得
2630	别墅	3130	高涨	3630	冷酷	4130	失踪	4630	一流
2631	别致	3131	稿件	3631	冷落	4131	师范	4631	一目了然
2632	别扭	3132	告辞	3632	冷却	4132	施加	4632	一如既往
2633	濒临	3133	告诫	3633	愣	4133	施展	4633	一丝不苟
2634	冰雹	3134	疙瘩	3634	黎明	4134	十足	4634	一向
2635	丙	3135	鸽子	3635	礼节	4135	石油	4635	衣裳
2636	并非	3136	搁	3636	礼尚往来	4136	时常	4636	依旧
2637	并列	3137	割	3637	里程碑	4137	时而	4637	依据
2638	拨	3138	歌颂	3638	理睬	4138	时光	4638	依靠
2639	波浪	3139	革命	3639	理所当然	4139	时机	4639	依赖
2640	波涛	3140	格局	3640	理直气壮	4140	时事	4640	依托
2641	剥削	3141	格式	3641	理智	4141	识别	4641	仪器
2642	播种	3142	隔阂	3642	力求	4142	实惠	4642	仪式
2643	伯母	3143	隔离	3643	力所能及	4143	实力	4643	遗产
2644	博大精深	3144	个体	3644	力争	4144	实施	4644	遗传
2645	博览会	3145	各抒己见	3645	历代	4145	实事求是	4645	遗留

2646	搏斗	3146	根深蒂固	3646	历来	4146	实行	4646	遗失
2647	薄弱	3147	根源	3647	立场	4147	实质	4647	疑惑
2648	补偿	3148	跟前	3648	立方	4148	拾	4648	以便
2649	补救	3149	跟随	3649	立交桥	4149	使命	4649	以免
2650	补贴	3150	跟踪	3650	立体	4150	示范	4650	以往
2651	捕捉	3151	更新	3651	立足	4151	示威	4651	以至
2652	哺乳	3152	更正	3652	利害	4152	示意	4652	以致
2653	不得已	3153	耕地	3653	例外	4153	世代	4653	亦
2654	不妨	3154	工艺品	3654	粒	4154	势必	4654	异常
2655	不敢当	3155	公安局	3655	连年	4155	势力	4655	意料
2656	不顾	3156	公道	3656	连锁	4156	事故	4656	意识
2657	不禁	3157	公告	3657	连同	4157	事迹	4657	意图
2658	不堪	3158	公关	3658	联欢	4158	事件	4658	意味着
2659	不可思议	3159	公民	3659	联络	4159	事态	4659	意向
2660	不愧	3160	公然	3660	联盟	4160	事务	4660	意志
2661	不料	3161	公认	3661	联想	4161	事项	4661	毅力
2662	不免	3162	公式	3662	廉洁	4162	事业	4662	毅然
2663	不时	3163	公务	3663	良心	4163	试图	4663	翼
2664	不惜	3164	公正	3664	谅解	4164	试验	4664	阴谋
2665	不相上下	3165	公证	3665	晾	4165	视力	4665	音响
2666	不像话	3166	功劳	3666	辽阔	4166	视频	4666	引导
2667	不屑一顾	3167	功效	3667	列举	4167	视线	4667	引擎
2668	不言而喻	3168	攻击	3668	临床	4168	视野	4668	引用
2669	不由得	3169	攻克	3669	淋	4169	是非	4669	饮食
2670	不择手段	3170	供不应求	3670	吝啬	4170	适宜	4670	隐蔽
2671	不止	3171	供给	3671	伶俐	4171	逝世	4671	隐患
2672	布告	3172	宫殿	3672	灵感	4172	释放	4672	隐瞒
2673	布局	3173	恭敬	3673	灵魂	4173	收藏	4673	隐私
2674	布置	3174	巩固	3674	灵敏	4174	收缩	4674	隐约
2675	步伐	3175	共和国	3675	凌晨	4175	收益	4675	英明
2676	部署	3176	共计	3676	零星	4176	收音机	4676	英勇
2677	部位	3177	共鸣	3677	领会	4177	手法	4677	婴儿
2678	才干	3178	勾结	3678	领事馆	4178	手势	4678	迎面
2679	财富	3179	钩子	3679	领土	4179	手艺	4679	盈利
2680	财务	3180	构思	3680	领悟	4180	守护	4680	应酬
2681	财政	3181	孤独	3681	领先	4181	首饰	4681	应邀
2682	裁缝	3182	孤立	3682	领袖	4182	首要	4682	拥护
2683	裁判	3183	姑且	3683	溜	4183	受罪	4683	拥有
2684	裁员	3184	辜负	3684	留恋	4184	授予	4684	庸俗
2685	采购	3185	古董	3685	留念	4185	书法	4685	永恒
2686	采集	3186	古怪	3686	留神	4186	书籍	4686	勇于

2687	采纳	3187	股东	3687	流浪	4187	书记	4687	涌现
2688	彩票	3188	股份	3688	流露	4188	书面	4688	踊跃
2689	参谋	3189	骨干	3689	流氓	4189	舒畅	4689	用户
2690	参照	3190	鼓动	3690	流通	4190	疏忽	4690	优胜劣汰
2691	残疾	3191	固然	3691	聋哑	4191	疏远	4691	优先
2692	残酷	3192	固体	3692	隆重	4192	束	4692	优异
2693	残留	3193	固有	3693	垄断	4193	束缚	4693	优越
2694	残忍	3194	固执	3694	笼罩	4194	树立	4694	忧郁
2695	灿烂	3195	故乡	3695	搂	4195	竖	4695	犹如
2696	仓促	3196	故障	3696	炉灶	4196	数额	4696	油腻
2697	仓库	3197	顾虑	3697	屡次	4197	耍	4697	油漆
2698	苍白	3198	顾问	3698	履行	4198	衰老	4698	有条不紊
2699	舱	3199	雇佣	3699	掠夺	4199	衰退	4699	幼稚
2700	操劳	3200	拐杖	3700	轮船	4200	率领	4700	诱惑
2701	操练	3201	关怀	3701	轮廓	4201	涮火锅	4701	渔民
2702	操纵	3202	关照	3702	轮胎	4202	双胞胎	4702	愚蠢
2703	操作	3203	观光	3703	论坛	4203	爽快	4703	愚昧
2704	嘈杂	3204	官方	3704	论证	4204	水利	4704	舆论
2705	草案	3205	管辖	3705	啰唆	4205	水龙头	4705	与日俱增
2706	草率	3206	贯彻	3706	络绎不绝	4206	水泥	4706	宇宙
2707	侧面	3207	惯例	3707	落成	4207	瞬间	4707	羽绒服
2708	测量	3208	灌溉	3708	落实	4208	司法	4708	玉
2709	策划	3209	罐	3709	麻痹	4209	司令	4709	预料
2710	策略	3210	光彩	3710	麻木	4210	私自	4710	预期
2711	层出不穷	3211	光辉	3711	麻醉	4211	思念	4711	预算
2712	层次	3212	光芒	3712	码头	4212	思索	4712	预先
2713	差别	3213	光荣	3713	蚂蚁	4213	思维	4713	预言
2714	插座	3214	广阔	3714	嘛	4214	斯文	4714	预兆
2715	查获	3215	归根到底	3715	埋伏	4215	死亡	4715	欲望
2716	岔	3216	归还	3716	埋没	4216	四肢	4716	寓言
2717	刹那	3217	规范	3717	埋葬	4217	寺庙	4717	愈
2718	诧异	3218	规格	3718	迈	4218	饲养	4718	冤枉
2719	柴油	3219	规划	3719	脉搏	4219	肆无忌惮	4719	元首
2720	搀	3220	规章	3720	埋怨	4220	耸	4720	元素
2721	馋	3221	轨道	3721	蔓延	4221	艘	4721	元宵节
2722	缠绕	3222	贵族	3722	漫长	4222	苏醒	4722	园林
2723	产业	3223	跪	3723	漫画	4223	俗话	4723	原告
2724	阐述	3224	棍棒	3724	慢性	4224	诉讼	4724	原理
2725	颤抖	3225	国防	3725	忙碌	4225	素食	4725	原始
2726	昌盛	3226	国务院	3726	盲目	4226	素质	4726	原先
2727	尝试	3227	果断	3727	茫茫	4227	塑造	4727	圆满

2728	偿还	3228	过度	3728	茫然	4228	算数	4728	缘故
2729	场合	3229	过渡	3729	茂盛	4229	随即	4729	源泉
2730	场面	3230	过奖	3730	冒充	4230	随意	4730	约束
2731	场所	3231	过滤	3731	冒犯	4231	岁月	4731	乐谱
2732	敞开	3232	过失	3732	枚	4232	隧道	4732	岳母
2733	畅通	3233	过问	3733	媒介	4233	损坏	4733	孕育
2734	畅销	3234	过瘾	3734	美观	4234	索取	4734	运算
2735	倡导	3235	过于	3735	美满	4235	索性	4735	运行
2736	倡议	3236	嗨	3736	美妙	4236	塌	4736	酝酿
2737	钞票	3237	海拔	3737	萌芽	4237	踏实	4737	蕴藏
2738	超越	3238	海滨	3738	猛烈	4238	塔	4738	熨
2739	巢穴	3239	含糊	3739	眯	4239	台风	4739	杂技
2740	朝代	3240	含义	3740	弥补	4240	太空	4740	杂交
2741	嘲笑	3241	寒暄	3741	弥漫	4241	泰斗	4741	砸
2742	潮流	3242	罕见	3742	迷惑	4242	贪婪	4742	咋
2743	撤退	3243	捍卫	3743	迷人	4243	贪污	4743	灾难
2744	撤销	3244	行列	3744	迷信	4244	摊	4744	栽培
2745	沉淀	3245	航空	3745	谜语	4245	瘫痪	4745	宰
2746	沉闷	3246	航天	3746	密度	4246	弹性	4746	再接再厉
2747	沉思	3247	航行	3747	密封	4247	坦白	4747	在意
2748	沉重	3248	毫米	3748	棉花	4248	叹气	4748	攒
2749	沉着	3249	毫无	3749	免得	4249	探测	4749	暂且
2750	陈旧	3250	豪迈	3750	免疫	4250	探索	4750	赞叹
2751	陈列	3251	号召	3751	勉励	4251	探讨	4751	赞助
2752	陈述	3252	耗费	3752	勉强	4252	探望	4752	遭受
2753	衬托	3253	呵	3753	面貌	4253	倘若	4753	遭殃
2754	称心如意	3254	合并	3754	面子	4254	掏	4754	遭遇
2755	称号	3255	合成	3755	描绘	4255	滔滔不绝	4755	糟蹋
2756	成本	3256	合伙	3756	瞄准	4256	陶瓷	4756	造型
2757	成交	3257	合算	3757	渺小	4257	陶醉	4757	噪音
2758	成天	3258	和蔼	3758	藐视	4258	淘汰	4758	责怪
2759	成效	3259	和解	3759	灭亡	4259	讨好	4759	贼
2760	成心	3260	和睦	3760	蔑视	4260	特长	4760	增添
2761	成员	3261	和气	3761	民间	4261	特定	4761	赠送
2762	呈现	3262	和谐	3762	民主	4262	特意	4762	扎
2763	诚挚	3263	嘿	3763	敏捷	4263	提拔	4763	扎实
2764	承办	3264	痕迹	3764	敏锐	4264	提炼	4764	渣
2765	承包	3265	狠心	3765	名次	4265	提示	4765	眨
2766	承诺	3266	恨不得	3766	名额	4266	提议	4766	诈骗
2767	城堡	3267	横	3767	名副其实	4267	题材	4767	摘要
2768	乘	3268	哼	3768	名誉	4268	体裁	4768	债券

2769	盛	3269	轰动	3769	明明	4269	体积	4769	沾光
2770	惩罚	3270	烘	3770	明智	4270	体谅	4770	瞻仰
2771	澄清	3271	宏观	3771	命名	4271	体面	4771	斩钉截铁
2772	橙	3272	宏伟	3772	摸索	4272	体系	4772	展示
2773	秤	3273	洪水	3773	模范	4273	天才	4773	展望
2774	吃苦	3274	哄	3774	模式	4274	天赋	4774	展现
2775	吃力	3275	喉咙	3775	模型	4275	天伦之乐	4775	崭新
2776	迟钝	3276	吼	3776	膜	4276	天然气	4776	占据
2777	迟缓	3277	后代	3777	摩擦	4277	天生	4777	占领
2778	迟疑	3278	后顾之忧	3778	磨合	4278	天堂	4778	战斗
2779	持久	3279	后勤	3779	魔鬼	4279	天文	4779	战略
2780	赤道	3280	候选	3780	魔术	4280	田径	4780	战术
2781	赤字	3281	呼唤	3781	抹杀	4281	田野	4781	战役
2782	冲动	3282	呼啸	3782	莫名其妙	4282	舔	4782	章程
2783	冲击	3283	呼吁	3783	墨水儿	4283	挑剔	4783	帐篷
2784	冲突	3284	忽略	3784	默默	4284	条款	4784	障碍
2785	充当	3285	胡乱	3785	谋求	4285	条理	4785	招标
2786	充沛	3286	胡须	3786	模样	4286	条约	4786	招收
2787	充实	3287	湖泊	3787	母语	4287	调和	4787	朝气蓬勃
2788	充足	3288	花瓣	3788	目睹	4288	调剂	4788	着迷
2789	重叠	3289	花蕾	3789	目光	4289	调节	4789	沼泽
2790	崇拜	3290	华丽	3790	沐浴	4290	调解	4790	照样
2791	崇高	3291	华侨	3791	拿手	4291	调料	4791	照耀
2792	崇敬	3292	化肥	3792	纳闷儿	4292	挑拨	4792	折腾
2793	稠密	3293	化石	3793	耐用	4293	挑衅	4793	遮挡
2794	筹备	3294	化验	3794	南辕北辙	4294	跳跃	4794	折
2795	丑恶	3295	化妆	3795	难得	4295	亭子	4795	折磨
2796	出路	3296	划分	3796	难堪	4296	停泊	4796	侦探
2797	出卖	3297	画蛇添足	3797	难能可贵	4297	停顿	4797	珍贵
2798	出身	3298	话筒	3798	恼火	4298	停滞	4798	珍稀
2799	出神	3299	欢乐	3799	内涵	4299	挺拔	4799	珍珠
2800	出息	3300	还原	3800	内幕	4300	通货膨胀	4800	真理
2801	初步	3301	环节	3801	内在	4301	通缉	4801	真相
2802	除	3302	缓和	3802	能量	4302	通俗	4802	真挚
2803	处分	3303	患者	3803	拟定	4303	通讯	4803	斟酌
2804	处境	3304	荒凉	3804	逆行	4304	通用	4804	枕头
2805	处置	3305	荒谬	3805	年度	4305	同胞	4805	阵地
2806	储备	3306	荒唐	3806	捏	4306	同志	4806	阵容
2807	储存	3307	皇帝	3807	凝固	4307	铜	4807	振奋
2808	储蓄	3308	皇后	3808	凝聚	4308	童话	4808	振兴
2809	触犯	3309	黄昏	3809	凝视	4309	统筹兼顾	4809	震撼

2810	川流不息	3310	恍然大悟	3810	拧	4310	统计	4810	震惊
2811	穿越	3311	晃	3811	宁肯	4311	统统	4811	镇定
2812	传达	3312	挥霍	3812	宁愿	4312	统治	4812	镇静
2813	传单	3313	辉煌	3813	扭转	4313	投机	4813	正月
2814	传授	3314	回报	3814	纽扣儿	4314	投票	4814	争端
2815	船舶	3315	回避	3815	农历	4315	投诉	4815	争夺
2816	喘气	3316	回顾	3816	浓厚	4316	投降	4816	争气
2817	串	3317	回收	3817	奴隶	4317	投掷	4817	争先恐后
2818	床单	3318	悔恨	3818	虐待	4318	透露	4818	争议
2819	创立	3319	毁灭	3819	挪	4319	秃	4819	征服
2820	创新	3320	汇报	3820	哦	4320	突破	4820	征收
2821	创业	3321	会晤	3821	殴打	4321	图案	4821	挣扎
2822	创作	3322	贿赂	3822	呕吐	4322	徒弟	4822	蒸发
2823	吹牛	3323	昏迷	3823	偶像	4323	途径	4823	整顿
2824	吹捧	3324	荤	3824	趴	4324	涂抹	4824	正当
2825	炊烟	3325	浑身	3825	排斥	4325	土壤	4825	正负
2826	垂直	3326	混合	3826	排除	4326	团结	4826	正规
2827	锤	3327	混乱	3827	排放	4327	团体	4827	正经
2828	纯粹	3328	混淆	3828	排练	4328	团圆	4828	正气
2829	纯洁	3329	混浊	3829	徘徊	4329	推测	4829	正义
2830	慈善	3330	活该	3830	派别	4330	推翻	4830	正宗
2831	慈祥	3331	活力	3831	派遣	4331	推理	4831	证实
2832	磁带	3332	火箭	3832	攀登	4332	推论	4832	证书
2833	雌雄	3333	火焰	3833	盘旋	4333	推销	4833	郑重
2834	次品	3334	火药	3834	判决	4334	吞吞吐吐	4834	政策
2835	次序	3335	货币	3835	畔	4335	托运	4835	政权
2836	伺候	3336	讥笑	3836	庞大	4336	拖延	4836	症状
2837	刺	3337	饥饿	3837	抛弃	4337	脱离	4837	之际
2838	从容	3338	机动	3838	泡沫	4338	妥当	4838	支撑
2839	丛	3339	机构	3839	培育	4339	妥善	4839	支出
2840	凑合	3340	机灵	3840	配备	4340	妥协	4840	支流
2841	粗鲁	3341	机密	3841	配偶	4341	椭圆	4841	支配
2842	窜	3342	机械	3842	配套	4342	唾弃	4842	支援
2843	摧残	3343	机遇	3843	盆地	4343	挖掘	4843	支柱
2844	脆弱	3344	机智	3844	烹饪	4344	哇	4844	枝
2845	搓	3345	基地	3845	捧	4345	娃娃	4845	知觉
2846	磋商	3346	基金	3846	批发	4346	瓦解	4846	知足常乐
2847	挫折	3347	基因	3847	批判	4347	歪曲	4847	脂肪
2848	搭	3348	激发	3848	劈	4348	外表	4848	执行
2849	搭档	3349	激励	3849	皮革	4349	外行	4849	执着
2850	搭配	3350	激情	3850	疲惫	4350	外界	4850	直播

2851	达成	3351	及早	3851	疲倦	4351	外向	4851	直径
2852	答辩	3352	吉祥	3852	屁股	4352	丸	4852	侄子
2853	答复	3353	级别	3853	譬如	4353	完备	4853	值班
2854	打包	3354	极端	3854	偏差	4354	完毕	4854	职能
2855	打官司	3355	极限	3855	偏见	4355	玩弄	4855	职位
2856	打击	3356	即便	3856	偏僻	4356	玩意儿	4856	职务
2857	打架	3357	即将	3857	偏偏	4357	顽固	4857	殖民地
2858	打量	3358	急功近利	3858	片断	4358	顽强	4858	指标
2859	打猎	3359	急剧	3859	片刻	4359	挽回	4859	指定
2860	打仗	3360	急切	3860	漂浮	4360	挽救	4860	指甲
2861	大不了	3361	急于求成	3861	飘扬	4361	惋惜	4861	指令
2862	大臣	3362	急躁	3862	撇	4362	万分	4862	指南针
2863	大伙儿	3363	疾病	3863	拼搏	4363	往常	4863	指示
2864	大肆	3364	集团	3864	拼命	4364	往事	4864	指望
2865	大体	3365	嫉妒	3865	贫乏	4365	妄想	4865	指责
2866	大意	3366	籍贯	3866	贫困	4366	危机	4866	志气
2867	大致	3367	给予	3867	频繁	4367	威风	4867	制裁
2868	歹徒	3368	计较	3868	频率	4368	威力	4868	制服
2869	代价	3369	记性	3869	品尝	4369	威望	4869	制约
2870	代理	3370	记载	3870	品德	4370	威信	4870	制止
2871	带领	3371	纪要	3871	品质	4371	微不足道	4871	治安
2872	怠慢	3372	技巧	3872	品种	4372	微观	4872	治理
2873	逮捕	3373	忌讳	3873	平凡	4373	为难	4873	致辞
2874	担保	3374	季度	3874	平面	4374	为期	4874	致力
2875	胆怯	3375	季军	3875	平坦	4375	违背	4875	致使
2876	诞辰	3376	迹象	3876	平行	4376	唯独	4876	智力
2877	诞生	3377	继承	3877	平庸	4377	维持	4877	智能
2878	淡季	3378	寄托	3878	平原	4378	维护	4878	智商
2879	淡水	3379	寂静	3879	评估	4379	维生素	4879	滞留
2880	蛋白质	3380	加工	3880	评论	4380	伪造	4880	中断
2881	当场	3381	加剧	3881	屏幕	4381	委托	4881	中立
2882	当初	3382	夹杂	3882	屏障	4382	委员	4882	中央
2883	当代	3383	佳肴	3883	坡	4383	卫星	4883	忠诚
2884	当面	3384	家常	3884	泼	4384	未免	4884	忠实
2885	当前	3385	家伙	3885	颇	4385	畏惧	4885	终点
2886	当事人	3386	家属	3886	迫不及待	4386	喂（动词）	4886	终究
2887	当务之急	3387	家喻户晓	3887	迫害	4387	蔚蓝	4887	终身
2888	当选	3388	尖端	3888	破例	4388	慰问	4888	终止
2889	党	3389	尖锐	3889	魄力	4389	温带	4889	衷心
2890	档案	3390	坚定	3890	扑	4390	温和	4890	肿瘤
2891	档次	3391	坚固	3891	铺	4391	文凭	4891	种子

2892	导弹	3392	坚韧	3892	朴实	4392	文物	4892	种族
2893	导航	3393	坚实	3893	朴素	4393	文献	4893	众所周知
2894	导向	3394	坚硬	3894	普及	4394	文雅	4894	种植
2895	捣乱	3395	艰难	3895	瀑布	4395	文艺	4895	重心
2896	倒闭	3396	监督	3896	凄凉	4396	问世	4896	舟
2897	盗窃	3397	监视	3897	期望	4397	窝	4897	州
2898	稻谷	3398	监狱	3898	期限	4398	乌黑	4898	周边
2899	得不偿失	3399	煎	3899	欺负	4399	污蔑	4899	周密
2900	得力	3400	拣	3900	欺骗	4400	诬陷	4900	周年
2901	得天独厚	3401	检讨	3901	齐全	4401	无比	4901	周期
2902	得罪	3402	检验	3902	齐心协力	4402	无偿	4902	周折
2903	灯笼	3403	剪彩	3903	奇妙	4403	无耻	4903	周转
2904	登陆	3404	简化	3904	歧视	4404	无动于衷	4904	粥
2905	登录	3405	简陋	3905	旗袍	4405	无非	4905	昼夜
2906	蹬	3406	简体字	3906	旗帜	4406	无辜	4906	皱纹
2907	等候	3407	简要	3907	乞丐	4407	无精打采	4907	株
2908	等级	3408	见多识广	3908	岂有此理	4408	无赖	4908	诸位
2909	瞪	3409	见解	3909	企图	4409	无理取闹	4909	逐年
2910	堤坝	3410	见闻	3910	启程	4410	无能为力	4910	主办
2911	敌视	3411	见义勇为	3911	启蒙	4411	无穷无尽	4911	主导
2912	抵达	3412	间谍	3912	启示	4412	无微不至	4912	主管
2913	抵抗	3413	间隔	3913	启事	4413	无忧无虑	4913	主流
2914	抵制	3414	间接	3914	起草	4414	无知	4914	主权
2915	地步	3415	剑	3915	起初	4415	武器	4915	主义
2916	地势	3416	健全	3916	起伏	4416	武侠	4916	拄
2917	地质	3417	舰艇	3917	起哄	4417	武装	4917	嘱咐
2918	递增	3418	践踏	3918	起码	4418	侮辱	4918	助理
2919	颠簸	3419	溅	3919	起源	4419	舞蹈	4919	助手
2920	颠倒	3420	鉴别	3920	气概	4420	务必	4920	住宅
2921	典礼	3421	鉴定	3921	气功	4421	物美价廉	4921	注射
2922	典型	3422	鉴于	3922	气魄	4422	物业	4922	注视
2923	点缀	3423	将近	3923	气色	4423	物资	4923	注释
2924	电源	3424	将就	3924	气势	4424	误差	4924	注重
2925	垫	3425	将军	3925	气味	4425	误解	4925	驻扎
2926	惦记	3426	僵硬	3926	气象	4426	夕阳	4926	著作
2927	奠定	3427	奖励	3927	气压	4427	昔日	4927	铸造
2928	叼	3428	奖赏	3928	气质	4428	牺牲	4928	拽
2929	雕刻	3429	桨	3929	迄今为止	4429	溪	4929	专长
2930	雕塑	3430	降临	3930	器材	4430	熄灭	4930	专程
2931	吊	3431	交叉	3931	器官	4431	膝盖	4931	专利
2932	调动	3432	交代	3932	掐	4432	习俗	4932	专题

2933	跌	3433	交涉	3933	洽谈	4433	袭击	4933	砖
2934	丁	3434	交易	3934	恰当	4434	媳妇	4934	转达
2935	叮嘱	3435	娇气	3935	恰到好处	4435	喜闻乐见	4935	转让
2936	盯	3436	焦点	3936	恰巧	4436	喜悦	4936	转移
2937	定期	3437	焦急	3937	千方百计	4437	系列	4937	转折
2938	定义	3438	角落	3938	迁就	4438	细胞	4938	传记
2939	丢人	3439	侥幸	3939	迁徙	4439	细菌	4939	庄稼
2940	丢三落四	3440	搅拌	3940	牵	4440	细致	4940	庄严
2941	东道主	3441	缴纳	3941	牵扯	4441	峡谷	4941	庄重
2942	东张西望	3442	较量	3942	牵制	4442	狭隘	4942	装备
2943	董事长	3443	教养	3943	谦逊	4443	狭窄	4943	装卸
2944	动荡	3444	阶层	3944	签署	4444	霞	4944	壮观
2945	动机	3445	皆	3945	前景	4445	下属	4945	壮丽
2946	动静	3446	接连	3946	前提	4446	先进	4946	壮烈
2947	动力	3447	揭露	3947	潜力	4447	先前	4947	幢
2948	动脉	3448	节制	3948	潜水	4448	纤维	4948	追悼
2949	动身	3449	节奏	3949	潜移默化	4449	掀起	4949	追究
2950	动手	3450	杰出	3950	谴责	4450	鲜明	4950	坠
2951	动态	3451	结晶	3951	强制	4451	闲话	4951	准则
2952	动员	3452	结局	3952	抢劫	4452	贤惠	4952	卓越
2953	冻结	3453	结算	3953	抢救	4453	弦	4953	着手
2954	栋	3454	截止	3954	强迫	4454	衔接	4954	着想
2955	兜	3455	截至	3955	桥梁	4455	嫌	4955	着重
2956	陡峭	3456	竭尽全力	3956	窍门	4456	嫌疑	4956	琢磨
2957	斗争	3457	解除	3957	翘	4457	显著	4957	姿态
2958	督促	3458	解放	3958	切实	4458	现场	4958	资本
2959	毒品	3459	解雇	3959	锲而不舍	4459	现成	4959	资产
2960	独裁	3460	解剖	3960	钦佩	4460	现状	4960	资深
2961	堵塞	3461	解散	3961	侵犯	4461	线索	4961	资助
2962	赌博	3462	解体	3962	侵略	4462	宪法	4962	滋润
2963	杜绝	3463	戒备	3963	亲密	4463	陷害	4963	滋味
2964	端	3464	界限	3964	亲热	4464	陷阱	4964	子弹
2965	端午节	3465	借鉴	3965	勤俭	4465	陷入	4965	自卑
2966	端正	3466	借助	3966	勤劳	4466	馅儿	4966	自发
2967	短促	3467	金融	3967	倾听	4467	乡镇	4967	自力更生
2968	断定	3468	津津有味	3968	倾向	4468	相差	4968	自满
2969	断绝	3469	紧迫	3969	倾斜	4469	相等	4969	自主
2970	堆积	3470	锦上添花	3970	清澈	4470	相辅相成	4970	宗教
2971	队伍	3471	进而	3971	清晨	4471	相应	4971	宗旨
2972	对策	3472	进攻	3972	清除	4472	镶嵌	4972	棕色
2973	对称	3473	进化	3973	清洁	4473	响亮	4973	踪迹

2974	对付	3474	进展	3974	清理	4474	响应	4974	总而言之
2975	对抗	3475	近来	3975	清晰	4475	想方设法	4975	总和
2976	对立	3476	晋升	3976	清醒	4476	向导	4976	纵横
2977	对联	3477	浸泡	3977	清真	4477	向来	4977	走廊
2978	对应	3478	茎	3978	情报	4478	向往	4978	走漏
2979	对照	3479	经费	3979	情节	4479	巷	4979	走私
2980	兑现	3480	经纬	3980	情理	4480	相声	4980	揍
2981	顿时	3481	惊动	3981	情形	4481	削	4981	租赁
2982	多元化	3482	惊奇	3982	晴朗	4482	消除	4982	足以
2983	哆嗦	3483	惊讶	3983	请柬	4483	消毒	4983	阻碍
2984	堕落	3484	兢兢业业	3984	请教	4484	消防	4984	阻拦
2985	额外	3485	精打细算	3985	请示	4485	消耗	4985	阻挠
2986	恶心	3486	精华	3986	请帖	4486	消灭	4986	祖父
2987	恶化	3487	精简	3987	丘陵	4487	销毁	4987	祖国
2988	遏制	3488	精密	3988	区分	4488	潇洒	4988	祖先
2989	恩怨	3489	精确	3989	区域	4489	小心翼翼	4989	钻研
2990	而已	3490	精通	3990	曲折	4490	肖像	4990	钻石
2991	二氧化碳	3491	精心	3991	驱逐	4491	效益	4991	嘴唇
2992	发布	3492	精益求精	3992	屈服	4492	协会	4992	罪犯
2993	发财	3493	精致	3993	渠道	4493	协商	4993	尊严
2994	发呆	3494	井	3994	曲子	4494	协调	4994	遵循
2995	发动	3495	颈椎	3995	取缔	4495	协议	4995	作弊
2996	发觉	3496	警告	3996	趣味	4496	协助	4996	作废
2997	发射	3497	警惕	3997	圈套	4497	携带	4997	作风
2998	发誓	3498	竞赛	3998	权衡	4498	泄露	4998	作息
2999	发行	3499	竞选	3999	权威	4499	泄气	4999	座右铭
3000	发炎	3500	敬礼	4000	全局	4500	屑	5000	做主

謝辞

本論文の執筆にあたり、多くの方々にご指導ご支援を賜りました。この場を借りてお礼申し上げます。

まず、主指導教員である田畑智司先生には、大阪大学大学院言語文化研究科博士後期課程に進学した時から、3 年間にわたり、研究遂行と論文執筆において始終丁寧なご指導とご鞭撻をいただきました。田畑智司先生はコーパス言語学とデジタル・ヒューマニティーズをはじめ、人文科学研究とテキスト分析技術の面白さを私に教えてくださいました。心より深謝申し上げます。

次に、副指導教員である田中智行先生には、常に温かく励ましをいただきました。研究と将来について有益なご助言をくださったとともに、本論文の執筆にあたり、懇切丁寧かつ熱心なご指導とご教示をくださいました。深く御礼申し上げます。

本論文の審査を引き受けてくださった瀧田恵巳先生には、本論文について、貴重なご意見をいただきました。心より深謝致します。

そして、大阪大学外国語学部の郭修静先生、張恒悦先生、李佳先生には、本論文に使用した中国語作文のデータ収集に多大なるご協力とご支援をいただきました。心より感謝申し上げます。

また、日頃の研究・生活において、張朝陽さん、黄晨雯さん、藤田郁さん、山道知可さん、John Tran さん、朴香芸さんには常に温かい応援をいただきました。心より感謝いたします。

さらに、次世代挑戦的研究者育成プロジェクトより研究奨励費と研究費をいただき、学業と研究活動に専念することができました。誠にありがとうございました。

最後に、絶えず励ましてくださった厦門大学の朱宇先生、いつも温かい目で見守ってくださった家族と親友に深く感謝の意を表します。