



副詞の文体の計量について——中俣論文の主成分分析結果との比較——

メタデータ	言語: Japanese 出版者: 北海道教育大学語学文学会 公開日: 2021-12-09 キーワード: 作成者: 馬場, 俊臣 メールアドレス: 所属:
URL	https://doi.org/10.32150/00010719

副詞の文体の計量について

—— 中俣論文の主成分分析結果との比較 ——

馬 場 俊 臣

1. はじめに

「文体」という用語は様々な意味で用いられる。『日本語 文章・文体・表現事典』の「文体」の項目¹によれば、「多様な文体観」は「文章の表現上の性格を他と対比的にとらえた特殊性を問題にしている」点ではすべてに共通しており、大別すると、「類型面」で捉えた「文体」（文語体と口語体、デアル体とデスマス体など）と「個性面」で捉えた「文体」（白樺派の文体、太宰治の文体など）とに二分できる。このように、「文体」は「文章」についてまず問題とされるが、文章の文体を決定する一つの要素として単語（語）の文体的特徴がある（宮島1977：871）。

語の文体については、「文体的特徴からする単語の分類は、連続的であり、程度の差によるものである。」（宮島1977：873）と指摘されている。

こうした「連続的であり、程度の差」のある語の文体差を数値化する試みの一つとして、馬場（2018a）で提案した専門度平均値、客観度平均値、硬度平均値、くだけ度平均値、語りかけ性度平均値（以下、これら5種類の平均値をまとめて「文体平均値」と仮称する。）がある。文体平均値は、国立国語研究所（2015）『BCCWJ 図書館サブコーパスの文体情報』を利用して算出した語の文体差を表す数値である。

文体平均値の性質と有効性を明らかにするために、馬場（2018a）、馬場（2018b）、馬場（2020）では、それぞれ、柏野（2016）の文体4段階²及び井上（2013）の「アンケート文体値」³との比較、柏野（2016）・田中（1987）・柴田（2017）の接続詞の文体的違い⁴との比較、佐野（2009、2016）の「語彙密度平均値」⁵との比較を行っている。これらの比較から、専門度平均値、客観度平均値、硬度平均値、くだけ度

¹ 中村（2011）。

² 主に接続詞と副詞を対象として、「書き言葉的」「話し言葉的」な違いによって、使用の目安を4段階に分類したものの。

³ 調査対象の語について、「レベル1 卑俗体」「レベル2 口頭体」「レベル3 汎用体」「レベル4 書記体」「レベル5 文章体」の5分類のいずれであるかの判断を求めたアンケート調査の結果に基づいて、文体の程度を数値化したもの。

⁴ 柏野（2016）については脚注2参照。田中（1987）では、接続詞を「会話的」「話ことば的」「一般」「書きことば的」「文語的」の5種類に分類している。柴田（2017）では、『日本語話し言葉コーパス』（CSJ）を利用して接続詞の文体差を点数化している。

⁵ 『現代日本語書き言葉均衡コーパス』（BCCWJ）の出版サブコーパス（図書）に含まれるサンプルの語彙密度のデータに基づいて算出した、語ごとの「書き言葉らしさ」「話し言葉らしさ」を表す数値。

平均値の4種類は、書き言葉のか話し言葉のかという観点⁶に基づく語の分類や数値と極めて相関が強いことが明らかとなっている。

本稿で取り上げる中俣(2020)は、副詞の文体差を計量的に分析した研究である。中俣(2020)では、副詞を対象とし、『現代日本語書き言葉均衡コーパス』(BCCWJ)⁷のレジスター別の調整頻度のデータを利用して主成分分析を行い、各副詞の第一主成分(PC1)及び第二主成分(PC2)の主成分得点に基づいて、対象とした副詞を3種4類に分類している。この分類で用いている主成分得点は連続的な数値であり、各副詞の何らかの文体的特徴を連続的な数値で表したものと見ることができる。

本稿では、文体平均値の性質と有効性をより明らかにするために、文体平均値と中俣(2020)のPC1及びPC2の主成分得点との関係を比較調査し分析考察する。

2. 文体平均値について

本節では、文体平均値について説明する。

文体平均値は、国立国語研究所(2015)『BCCWJ 図書館サブコーパスの文体情報』を利用して、語の文体差を「連続的な」数値として示したものである。

『BCCWJ 図書館サブコーパスの文体情報』⁸は、文体研究を進めるために、『現代日本語書き言葉均衡コーパス』(BCCWJ)の図書館サブコーパス収録のサンプルすべてに文体情報のアノテーションを行った成果を公開用にまとめたデータである。図書館サブコーパスの各サンプルに、「内容・表現の文体的特徴を表す分類指標」及び「形式・内容・表現に文体判断が単純にいかない特徴をもつものの分類指標」を付与している。専門度、客観度、硬度、くだけ度、語りかけ性度の5種類の分類指標は、「内容・表現の文体的特徴を表す分類指標」として設計されたものであり、それぞれ次の3段階～5段階のいずれかの段階が、各サンプルに付与されている⁹。

(a) 専門度 1 専門家向き/2 やや専門的な一般向き/3 一般向き/4 中高

⁶ 本稿では、「書き言葉の一話し言葉的」対立に、「硬い—柔らかい」や「改まった—くだけた」という文体的対立も含めている。なお、小磯(2012)では、テキストの多様性を捉える分類指標策定のために、テキストを読んだ際に感じる印象について12の尺度に基づき評定実験を行ったうえで、因子分析を行い、抽出された三つの因子を「スタイル」「抑揚・リズム性」「整然性」と解釈し、三つの分類指標と位置付けている。「スタイル」と解釈された因子は、「硬い—柔らかい」「改まった—くだけた」「書きことば的一話しことば的」の尺度が、高い正の因子負荷量を示している。この結果は、硬さ(硬度)、改まり(くだけ度)、書き言葉的・話し言葉的対立の3種類はテキストの印象として共通性が高いことを示すものである。

⁷ 国立国語研究所「概要 現代日本語書き言葉均衡コーパス(BCCWJ)」(<https://ccd.ninjal.ac.jp/bccwj/index.html>)参照。(2021年8月30日最終閲覧)

⁸ 説明は、国立国語研究所(2015)の添付文書「BCCWJ 図書館サブコーパスの文体情報説明書」及び柏野(2013)に基づく。

⁹ 付与は、言語データ構築経験有のおおよそ20～50代の女性、延べ9名の作業者によって行われている。

生向き／5 小学生・幼児向き

- (b) 客観度 1 とても客観的／2 どちらかといえば客観的／3 どちらかといえば主観的／4 とても主観的
- (c) 硬度 1 とても硬い／2 どちらかといえば硬い／3 どちらかといえば軟らかい／4 とても軟らかい
- (d) くだけ度 1 とてもくだけている／2 どちらかといえばくだけている／3 くだけていない
- (e) 語りかけ性度 1 とても語りかけ性がある／2 どちらかといえば語りかけ性がある／3 特に語りかけ性はない

文体平均値は、概略、次の手順によって算出される。硬度平均値を例にして説明する。図書館サブコーパス収録のサンプルすべてを対象として、ある短単位又は長単位¹⁰の語彙素（語）（ W_1 ）が出現するサンプルを特定する¹¹。出現するそれぞれのサンプルの硬度の段階の数字を『BCCWJ 図書館サブコーパスの文体情報』から得て、それを数値と見做して算術平均値を算出し、その値を W_1 の硬度平均値とする。ただし、 W_1 が同一のサンプルに複数回出現する場合は、そのサンプルでの出現は1回と見做して算術平均値を算出する。例えば、ある語（ W_1 ）が、硬度1のサンプルAで1回、硬度2のサンプルBで3回、硬度3のサンプルCで4回、それぞれ用いられていたとすると、 W_1 の硬度平均値は2となる¹²。より「硬い」サンプルにより多く使われている語は硬度平均値が低くなり、逆に、より「柔らかい」サンプルにより多く使われている語は、硬度平均値が高くなる。硬度平均値の高低は、柔らかさ・硬さという文体の違いであるとともに、後述するように、話し言葉のか書き言葉のかという文体の違いと見ることができる。

なお、馬場（2018a, 2018b）では語の単位として長単位の語彙素を用いた調査を行い、馬場（2020）では短単位の語彙素を用いた調査を行っている。本稿の調査では短単位の語彙素を用いて算出した文体平均値を用いる。

3. 中俣（2020）の概要

中俣（2020）は、日本語教育への応用を目指し、計量的手法を用いて、文体の観

¹⁰ BCCWJでは、語の単位として「短単位」及び「長単位」の2種類の単位を設けている。「短単位」は「用例収集に適した」単位であり「形態論的に一貫した言語単位」として認定されている単位である。「長単位」は「複合語を把握する」ことができ「サンプルの言語的特徴の解明に適した」単位である。また、「語彙素」は「国語辞典の見出し」に相当する。

（国立国語研究所コーパス開発センター（2015）参照）

¹¹ BCCWJのDVD版公開データ（BCCWJ-DVD版Version1.1）の短単位データ及び長単位データ（「DISC2_NT」（NumTrans版）の「TSV_SUW_NT」及び「TSV_LUW_NT」の各「LB.zip」）を用いている。

¹² $(1 + 2 + 3) \div 3 = 2$ 。

点から副詞の分類を試みたものである。「基本的な副詞」49語と「それと同様の機能を持つ文体的に異なる副詞」115語の副詞計164語を対象としている。まず、「副詞の文体的硬さ」を目的変数とし、説明変数として「BCCWJの異なるレジスター¹³における調整頻度」を利用して主成分分析を行っている。そして、各副詞における2つの主成分の主成分得点の値に基づいて、副詞を分類している。

具体的には、特に、説明力の高い2つの主成分（PC：Principal Component）について検討している。第一主成分（PC1）では、双方向的なコミュニケーションが保証されているレジスターである「Yahoo!知恵袋」「Yahoo!ブログ」が高い負の主成分負荷量を示し、「文学」を除く「出版・書籍」「図書館・書籍」が高い正の主成分負荷量を示していることに基づいて、PC1は「双方向的vs一方向的」の対立軸と捉えている。第二主成分（PC2）では、「文学」次いで（「文学」を除く）「ベストセラー」が高い正の主成分負荷量を示し、公共性の高いレジスターである「国会会議録」「白書」「広報紙」が高い負の主成分負荷量を示していることに基づいて、PC2は「自己表現的vs公共的」の対立軸と捉えている。

次に、対象とした副詞を、それぞれのPC1（「双方向的vs一方向的」）とPC2（「自己表現的vs公共的」）の主成分得点の正負に基づいて、「A 一方向・自己表現」「B 一方向・公共」「C1 双方向・自己表現」「C2 双方向・公共」の4類に分類している。ただし、C1とC2は「両領域の副詞に顕著な特性の違いが見られないこと」や「レジスターの分類にPC2があまり寄与していない」ことから「日本語教育での応用の可能性を考えて、双方向的な領域は1つにまとめ、3種4類として考察を行う」としている。

「A 一方向・自己表現」には、「しばしば」「しだいに」などの副詞が含まれ、「A 自分について表現する副詞」と位置付けている。ここに含まれる副詞について、「これらの副詞がよく使われる言語活動は、話し言葉ではスピーチコンテストや動画作成などの準備された場面であり、書き言葉では意見文や経験文といった自分を含む内容について書く作文・エッセイである。書き言葉でも話し言葉でも、あまり日常的に使われるとは言い難い。」（：428）としている。

「B 一方向・公共」には、「おおむね」「ほぼ」などの副詞が含まれ、「B 客観的に表現する副詞」と位置付けている。ここに含まれる副詞について、「これらの副詞を使うのは大学などで求められるレポートなどである。典型的には書き言葉とい

¹³ BCCWJには13種類のレジスターがあるが、「図書館・書籍」「出版・書籍」「ベストセラー」に含まれるNDC分類番号9から始まるサンプルを「文学」として、別途一つのレジスターとしてまとめている。その結果、14種類のレジスターにおける調整頻度を利用している。なお、「副詞ごとの出現数の違いの影響」を大きく受けることを避けるために「各調整頻度をそれぞれの副詞の総語数で割り、副詞間の頻度についても調整も行った」（：422）としている。

えるが、非常に硬い報告のスタイルでは話し言葉で用いられるかもしれない。」(：429)としている。

「C1 双方向・自己表現」及び「C2 双方向・公共」には、それぞれ「まさか」「たまたま」などの副詞及び「よろしく」「どうぞ」などの副詞が含まれ、それぞれ「C1 より洗練されたやりとりで使う副詞」及び「C2 日常的なやりとりで使う基本の副詞」と位置付けている。ここに含まれる副詞について、「C1やC2の副詞が使われるのは典型的には会話であるが、メール、SNSへの書き込みなども双方向的なやりとりに含まれるといえる。」(：429)としている。

さらに、「副詞のグループの評価」として、類義の副詞はほぼ異なる文体のグループに属することを確認している。

4. 比較調査の方法と結果

まず、比較調査の対象とした副詞について説明する。

中俣(2020)は、副詞164語を対象として、コーパス検索アプリケーション「中納言」を用いてBCCWJの用例を収集しているが、その際に、「短単位検索」で「基本的には品詞と語彙素の両方を指定」して検索を行っている。

文体平均値は、短単位の語彙素を対象とした値と長単位の語彙素を対象とした値があるが、中俣(2020)に合わせて短単位の語彙素を対象とした文体平均値を用いる¹⁴。なお、中俣(2020)では「それほど」「あつというま」のような2つ以上の短単位の語彙素による複合語も対象に含まれるが、全体としては1つの短単位ではないため対応する文体平均値は得られていない。そのため、このような複合語は比較調査の対象から除外する。また、「いま」「たしょう」のように短単位では「副詞」とされていない語も比較調査の対象から除外する¹⁵。

中俣(2020)で対象とした副詞164語のうち、品詞(副詞)と語彙素読み¹⁶が一致する101語¹⁷を比較調査の対象とする。この101語の文体平均値とPC1及びPC2の主成分得点のそれぞれとの散布図を描いたうえで、ピアソンの積率相関分析を行った。

図1は、左から、専門度平均値、客観度平均値、硬度平均値、くだけ度平均値、語りかけ性度平均値のそれぞれ(横軸)とPC1の主成分得点(縦軸)との散布図で

¹⁴ 長単位の語彙素の文体平均値を用いて、中俣(2020)で対象とした副詞と対応の取れる120語を対象として、同様の分析を行った。結果は、短単位の語彙素の文体平均値の場合とほぼ同様であった。

¹⁵ 「いま」「たしょう」などは、短単位での品詞は「名詞-普通名詞-副詞可能」である。なお、文体平均値の算出の際に文体指標の付与対象サンプル数が100未満となる語も除外している。「ある程度の大きさのサンプル数を確保するため」(馬場2018a：244)である。

¹⁶ 中俣(2020)の「別表 副詞の主成分得点とグループ」の各語の見出しは平仮名表記である。平仮名を片仮名に変換して、片仮名表記である「語彙素読み」と対応させた。

¹⁷ 比較調査の対象とした101語を、本稿末尾に示す。また、文体平均値とPC1及びPC2の主成分得点の例を本稿末尾の別表に示す。

ある。図2は、左から、専門度平均値、客観度平均値、硬度平均値、くだけ度平均値、語りかけ性度平均値のそれぞれ（横軸）とPC2の主成分得点（縦軸）との散布図である。

表1は、専門度平均値、客観度平均値、硬度平均値、くだけ度平均値、語りかけ性度平均値のそれぞれとPC1及びPC2の主成分得点との相関分析を行った結果である。

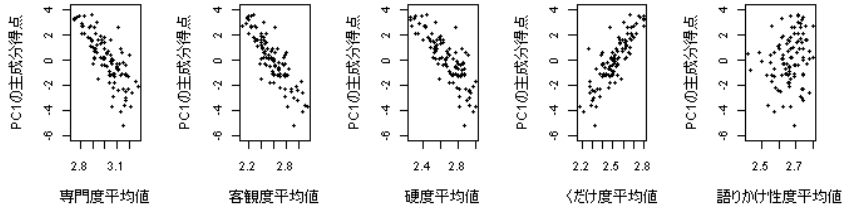


図1 文体平均値のそれぞれとPC1の主成分得点との散布図

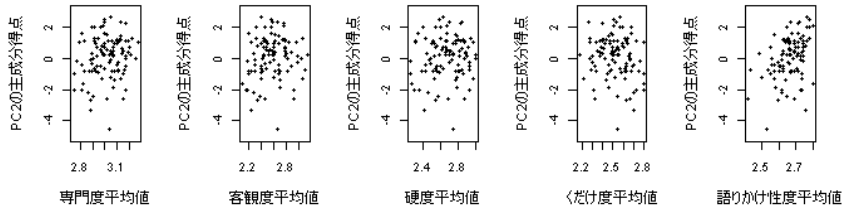


図2 文体平均値のそれぞれとPC2の主成分得点との散布図

表1 文体平均値のそれぞれとPC1及びPC2の主成分得点との相関係数

	専門度平均値	客観度平均値	硬度平均値	くだけ度平均値	語りかけ性度平均値
PC1の主成分得点	-0.785	-0.864	-0.833	0.857	0.211
PC2の主成分得点	0.184	0.128	0.100	-0.168	0.393

表1に示した通り、専門度平均値、客観度平均値、硬度平均値、くだけ度平均値のそれぞれとPC1の主成分得点との間には、強い（正又は負¹⁸）相関があることが示された¹⁹。それに対して、専門度平均値、客観度平均値、硬度平均値、くだけ

¹⁸ 専門度平均値、客観度平均値、硬度平均値については、それぞれ、「より専門的でない（すなわち、より低年齢向けである）」ほど、「より客観的でない（すなわち、より主観的である）」ほど、「より硬くない（すなわち、より柔らかい）」ほど、値が大きくなることに留意が必要である。

¹⁹ 無相関検定の結果、いずれも $p < .001$ であり、有意水準 5% において有意であった。

度平均値のそれぞれとPC2の主成分得点との間にはほぼ相関が見られなかった。また、語りかけ性度平均値とPC1及びPC2の主成分得点の間には弱い相関が見られた²⁰。

5. 考察

比較調査の結果に基づいて、考察を行う。

第一に、比較調査の結果、専門度平均値、客観度平均値、硬度平均値、くだけ度平均値のそれぞれとPC1の主成分得点の間には、強い（正又は負の）相関があることが分かった。

中俣（2020）では、PC1（第一主成分）に関して、「PC1では〈Yahoo!知恵袋〉〈Yahoo!ブログ〉が高い負の主成分負荷量を示し、〈出版・書籍*〉〈図書館・書籍*〉²¹が高い正の主成分負荷量を示すという関係になっている。両者はテナー²²の観点で、双方向的なやりとりを意識しているかどうかという違いがある。また、モード²³については全て書き言葉ではあるものの、話し言葉的な表現を多く使うかという点で違いがある。さらに、相手を意識することはその相手への配慮を生む。意識しなければ配慮もない。ここではこれらの対立軸をまとめて「双方向的vs一方向的」対立と捉える。」（：425）と述べている。また、「双方向性の強弱と同様の傾向は英語の文体研究でも見出されている。」（：426）として、Biber（1988）の文体の対立を表す「Involved versus Informational Production」の次元は「話し言葉か書き言葉かという媒体の違いを強く反映しているものである。本研究のPC1もこれに近い特性を表していると考えられる。」（：426）としている。

中俣（2020）のPC1は、「双方向的vs一方向的」対立と捉えられており、「話し言葉的対書き言葉的」という対立だけに帰することができない文体的対立ではあるが、上記の引用からも分かる通り、「話し言葉的対書き言葉的」という対立を内に含んでいることが分かる。

馬場（2018a）では、専門度平均値、客観度平均値、硬度平均値、くだけ度平均値の4種類の平均値は相互に強い相関があることを明らかにしている。さらに、これら4種類の平均値は、本稿2節でも述べたように、書き言葉的か話し言葉的かと

²⁰ 無相関検定の結果、PC1の主成分得点との相関分析では $p = .034$ 、PC2の主成分得点との相関分析では $p < .001$ であり、ともに有意水準5%において有意であった。

²¹ アスタリスク（*）の付いた「〈出版・書籍*〉〈図書館・書籍*〉」は、それぞれ、「出版・書籍」「図書館・書籍」のレジスターのサンプルから、NDC分類番号9から始まる「文学」のサンプルを除いたレジスターである。

²² 「テナーは役割関係とされ、「誰から誰に」なのかなど、参加者に関するパラメータである。」（中俣2020：425）

²³ 「モードは伝達様式とされ、書き言葉か話し言葉かといった媒体の違いのほか、テキストの地位や修辞様式も含む。」（中俣2020：425）

いう観点に基づく語の分類や数値と強い相関があることが明らかになっており、これら4種類の平均値は、書き言葉的か話し言葉的かの度合いを示す値である可能性が極めて高い。

専門度平均値、客観度平均値、硬度平均値、くだけ度平均値のそれぞれとPC1の主成分得点との間には強い相関があるという本稿の比較調査の結果は、上記のことを踏まえると、妥当な結果である。

さらに言えば、中俣（2020）のPC1は、「話し言葉的対書き言葉的」という対立だけに帰することができない対立であるにしても、専門度平均値、客観度平均値、硬度平均値、くだけ度平均値のそれぞれと強い相関があることから、「話し言葉的対書き言葉的」という対立の側面が強いと考えられる。

第二に、比較調査の結果、専門度平均値、客観度平均値、硬度平均値、くだけ度平均値のそれぞれとPC2（第二主成分）の主成分得点との間にはほぼ相関が見られなかった。中俣（2020）では、PC2は「自己表現的vs公共的」の対立軸と捉えており、本稿の比較調査の結果からも、PC2は「話し言葉的対書き言葉的」対立とは明らかに異なった文体的対立を反映しているものと考えられる。

第三に、比較調査の結果、語りかけ性度平均値は、PC1（第一主成分）及びPC2（第二主成分）の主成分得点のそれぞれの間と弱い相関が見られた。両者には、それぞれ何らかの関連性がある可能性もあるが、仮に関連性があつたとしても強いものではないであろう。各語の語りかけ性度平均値の性質は不明な部分が多く、弱い相関が見られた理由については、現段階では未詳である。

6. まとめ

本稿では、文体平均値の性質と有効性をより明らかにするために、文体平均値と中俣（2020）の主成分分析で得られた第一主成分（PC1）及び第二主成分（PC2）の主成分得点との関係を分析考察した。中俣（2020）では、副詞を対象として、BCCWJのレジスター別の調整頻度を用いて主成分分析を行い、その結果得られたPC1及びPC2の主成分得点を基準にして副詞の文体分類を行っている。中俣（2020）の対象とした副詞164語のうち、品詞（副詞）と語彙素読みが一致する101語（短単位の語彙素）を、本稿の比較調査の対象とし、この101語の文体平均値とPC1及びPC2の主成分得点のそれぞれとの散布図を描いたうえで、相関分析を行った。

相関分析の結果、専門度平均値、客観度平均値、硬度平均値、くだけ度平均値のそれぞれとPC1の主成分得点との間には強い相関があることが示された。専門度平均値、客観度平均値、硬度平均値、くだけ度平均値は、相互に強い相関があることは既に明らかになっており、これらの4種類の平均値は、書き言葉的か話し言葉的かの度合いを示す値である可能性が極めて高い。

中俣（2020）のPC1は、双方向的なやりとりを意識しているかどうか、話し言葉

的な表現を多く使うかどうか、相手を意識し相手への配慮をするかどうか、という対立があり、これらをまとめて「双方向的vs一方向的」対立と捉えている。本稿の分析結果は、中俣（2020）のPC1は、「話し言葉の対書き言葉的」という対立だけに帰することができない対立であるにしても、専門度平均値、客観度平均値、硬度平均値、くだけ度平均値のそれぞれと強い相関があることから、「話し言葉の対書き言葉的」という対立の側面が強いことを示している。

以上のように、本稿の分析結果は、文体平均値が語の文体差を表す目安の値として汎用性が高い可能性があることを強く示していると言える。

今後は、形容詞など他の品詞の語の文体差に関する先行研究の結果との比較も進めていく。また、語の文体平均値を利用して文章の文体の違いを推定する試みにも現在着手している。文体平均値の様々な活用方法を今後とも探っていきたい。

参考文献

- 井上次夫（2013）「単語の文体判断について（3）—話しことばと書きことば—」『全国大学国語教育学会 国語科教育研究 第125回広島大会研究発表要旨集』、全国大学国語教育学会、pp.303-306.
- 柏野和佳子（2013）「書籍サンプルの文体を分類する」『国語研プロジェクトレビュー』4（1）、国立国語研究所、pp.43-53.
- 柏野和佳子（2016）「学術的文章作成時に留意すべき「書き言葉的」「話し言葉的」な語の分類」『計量国語学会第六十回大会予稿集』、計量国語学会、pp.37-42.
- 小磯花絵（2012）「テキストの多様性をとらえる分類指標—体系化の試み—」石田基広、金明哲（編著）『コーパスとテキストマイニング』、共立出版、pp.70-82.
- 国立国語研究所（2015）『BCCWJ 図書館サブコーパスの文体情報』（2015年公開第1版）。（<http://doi.org/10.15084/00003109>）（2021年8月30日最終閲覧）
- 国立国語研究所コーパス開発センター（2015）『『現代日本語書き言葉均衡コーパス』利用の手引 第1.1版』。（<https://ccd.ninjal.ac.jp/bccwj/doc.html>）（2021年8月30日最終閲覧）
- 佐野大樹（2009）「話し言葉らしさ・書き言葉らしさ」の計測—語彙密度の日本語への適用性の検証—」『機能言語学研究』5、日本機能言語学会、pp.89-102.
- 佐野大樹（2016）「語彙密度から見た語彙シラバス」森篤嗣（編）『ニーズを踏まえた語彙シラバス』、くろしお出版、pp.79-93.
- 柴田好葉（2017）「話し言葉における接続詞の文体的特徴について」『言語資源活用ワークショップ発表論文集』2、国立国語研究所、pp.85-92.
- 田中章夫（1987）「接続の表現と語法（Ⅱ）」『日本語・日本文化』（14）、大阪外国語大学研究留学生別科、pp.1-19.（田中章夫（2001）『近代日本語の文法と表現』、明治書院 所収）

中俣尚己 (2020) 「主成分分析を用いた副詞の文体分析」『計量国語学』32 (7)、計量国語学会、pp.419-435.

中村明 (2011) 「文体」中村明、佐久間まゆみ、高崎みどり、十重田裕一、半沢幹一、宗像和重 (編) 『日本語 文章・文体・表現事典』、朝倉書店、p.126.

馬場俊臣 (2018a) 「『BCCWJ図書館サブコーパスの文体情報』を利用した語の文体差研究の可能性」『言語資源活用ワークショップ発表論文集』3、国立国語研究所、pp.241-256.

馬場俊臣 (2018b) 「接続詞の文体差の計量的分析の試み——『BCCWJ図書館サブコーパスの文体情報』を用いて——」『北海道教育大学紀要 人文科学・社会科学編』69 (1)、北海道教育大学、pp.1-14.

馬場俊臣 (2020) 「語の文体差を表す「文体値」について——「語彙密度平均値」と「硬度平均値」類との比較——」『札幌国語研究』(25)、北海道教育大学国語国文学会・札幌、pp.20-26.

宮島達夫 (1977) 「単語の文体的特徴」松村明教授還暦記念会 (編) 『松村明教授還暦記念 国語学と国語史』、明治書院、pp.871-903.

Biber, Douglas. (1988) *Variation across speech and writing*. Cambridge University Press.

資料 比較調査の対象とした101語 (語彙素)

敢えて、恰も、余り、改めて、如何に、幾ら、些か、一番、一層、一体、一杯、愈、色々、大いに、恐らく、凡そ、必ず、可成、急度、極めて、結構、極く、殊に、流石、更に、確り、屢、暫く、十分、随分、直ぐ、少し、ずっと、既に、精々、是非、全然、そう、相当、そろそろ、第一、大抵、大分、大変、沢山、直ちに、仮令、例えば、度々、多分、偶々、偶に、段々、丁度、一寸、どう、当然、どうぞ、到底、時折、特に、迎も、兎に角、兎も角、取り敢えず、取り分け、どンドン、猶、中々、成る可く、初めて、果たして、はっきり、再び、略、まあ、まさか、正に、先ず、益々、又、未だ、全く、丸で、無論、もう、若し、勿論、もっと、最も、最早、纏て、漸と、矢張り、やや、ゆっくり、漸く、良く、余程、より、宜しく

別表 比較調査のデータの例示 (PC1の主成分得点の上位及び下位各5語)

語彙素	PC1の主成分得点	PC2の主成分得点	専門度平均値	客観度平均値	硬度平均値	くだけ度平均値	語りかけ度平均値
凡そ	3.69	0.31	2.90	2.27	2.44	2.67	2.71
屢	3.59	1.04	2.79	2.24	2.31	2.80	2.71
最も	3.56	-0.79	2.83	2.21	2.38	2.75	2.67
取り分け	3.49	-0.16	2.77	2.19	2.25	2.75	2.74
より	3.33	-2.02	2.74	2.18	2.30	2.80	2.62
}							
偶に	-3.63	-0.74	3.16	3.10	2.94	2.28	2.54
まあ	-3.66	0.20	3.22	3.13	3.00	2.18	2.67
取り敢えず	-3.73	-0.86	3.05	2.82	2.82	2.34	2.62
結構	-4.08	-1.30	3.13	3.05	2.98	2.21	2.57
宜しく	-5.19	-2.63	3.16	2.92	2.87	2.33	2.69