文体指標と語彙の対応分析

浅原 正幸 †　加藤 祥 †　立花 幸子 †　柏野 和佳子 ‡†
（国立国語研究所　† コーパス開発センター　‡ 言語資源研究系）

Correspondence Analysis between Writing Styles and Lexicon

Masayuki Asahara, Sachi Kato, Sachiko Tachibana, and Wakako Kashino
(National Institute for Japanese Language and Linguistics)

要旨

柏野 (2013) は文体を分類するために設計した指標として，専門度，客観度，硬度，くだけ度，語りかけ性度の 5 種の分類指標を提案し，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 10,551 サンプルに対して悉皆的に付与を行った。本研究では，この分類指標に対して語彙素を特徴量とした対応分析を行い，各指標と特徴的な語彙素の分布の対応を品詞ごとに定量的に評価する。対応分析によってもたらされる第 1 主成分の寄与度に基づき，語彙素の分布のみによってとらえることが可能な指標とそうでない指標を明らかにする。さらに，語彙素の分布のみによってとらえられない指標については，どのような語彙素以外に利用すべき特徴量を検討する。

1. はじめに

コーパス調査において重要な要素として，利用するサンプルの文体情報がある。柏野 (2013), 柏野・奥村 (2012b) は文体を計量する指標として，専門度，客観度，硬度，くだけ度，語りかけ性度の 5 種の分類指標を提案し，『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 10,551 サンプルに対して悉皆的に付与を行った。このデータに対して，硬度・語りかけ性度を中心に，定量的・定性的な分析が進められてきた (柏野ほか (2012a), 保田ほか (2012b,a,c, 2013d,a,c,b), 加藤ほか (2015))。

本研究では統計的手法に基づいて，指標ごとに特徴的な語彙素を選定することを試みる。具体的には品詞ごとに高頻度語彙素を抽出し，指標に基づく対応分析を行い，語彙と指標との対応関係を分析する。この指標ごとに特徴的な語彙素の分布を確認し，語彙素のみにより表現可能な指標とそうでない指標を明らかにする。さらに，現在までに行われてきた定性的な分析と今回行った統計的な分析との比較を行う。

本研究の貢献は以下の通りである。各文体指標に対する人間の判断が，単純な語彙素の統計的な偏りにより表現可能な指標とそうでない指標を明らかにする。さらに語彙素の分布によってとらえられない指標については，語彙素以外に利用すべき特徴量を検討する。

2. 分析手法

2.1 文体指標

柏野 (2013) は文体指標として以下の 5 種類を規定した：

- 【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 中高生向き，5 小

学生・幼児向きの 5 段階指標

- 【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 とても主観的の 4 段階指標
- 【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 とても軟らかいの 4 段階指標
- 【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 くだけていないの 3 段階指標
- 【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の 3 段階指標

対象は BCCWJ に収録されている図書館サブコーパス 10,551 サンプル（書籍サンプル）とし，20〜50 代女性作業者延べ 9 名に可変長サンプルを呈示して文体指標付与を行った。作業において，インタビューなどのテキスト構造が文体付与に適さないものや外国語や数式などが多いサンプルなど内容や表現が文体付与に適さないものなど 1,664 サンプルを，文体指標付与対象から除外している。

2.2　対応分析

対応分析はクロス集計表の行と列の双方を並び替えることにより，行の項目と列の項目との相関関係を最大化するような処理を行う。基本的には主成分分析と同様にデータの分散を最大化する方向の軸（主成分）を逐次的に求め，説明変数を合成するという処理を行う。軸の選択は条件付極値問題として定式化でき，ラグランジュの未定乗数法によって解くと相関行列の固有値，固有ベクトルを求める問題に帰着する。値の大きい固有値に対応した軸から順に第 1 主成分，第 2 主成分と呼び，各軸は直交する。全固有値の総和で，各主成分に対応する固有値を割ったものを寄与率と呼び，各主成分によりどの程度説明ができているかの尺度となる。同様に第 1 主成分から第 α 主成分までの寄与率の和を第 α 主成分までの累積寄与率と呼び，当該主成分まででどの程度説明ができているかの尺度となる。

2.3　特徴量の設計

語彙素と品詞（細分類）の 2 つ組を 1 特徴量として設定する。先行研究で多く言及されている品詞大分類（動詞，連体詞，副詞，助動詞，助詞）ごとに頻度順に 30 語を抽出し特徴量とし，各指標に対して有効な語彙素と品詞を明らかにする。人手による指標の付与は可変長サンプルの地の文のみに対して行われた。対応分析は固定長サンプルと可変長サンプルの両方について行ったが，地の文と台詞の区別は行わず，全ての語彙について調査した。

固定長サンプルは文単位で 1000 字前後を選定するというランダムサンプリングに基づいたものであり，統計処理的にはより厳密なものである。一方，元の人手による指標の付与は文章構造上まとまった単位に行っており，実際のアノテーションにおいても文章構造に基づいた判定が行われている。

3.　結果（可変長サンプル）

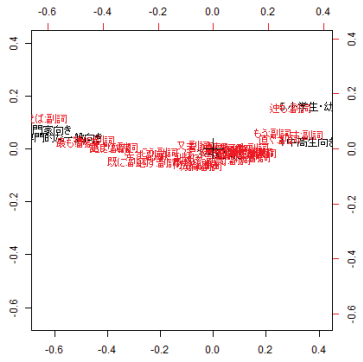
対応分析は固定長サンプルと可変長サンプルの両方について実施した。以下では可変長サンプルの分析結果について，指標ごとに特徴的な結果が得られた品詞や先行研究に多く言及され

ている品詞のみを示す。固定長サンプルの結果および可変長サンプルの他の品詞の結果については，第一著者に問い合わせていただければ提供する。

3.1　専門度

専門度については，ほぼすべての品詞において，統計的な手法によって得られた語彙と人手による判断との間に一致が見られた。以下では 5 品詞の中でもっとも第 1 主成分の寄与度が高かった副詞について言及する。

副詞



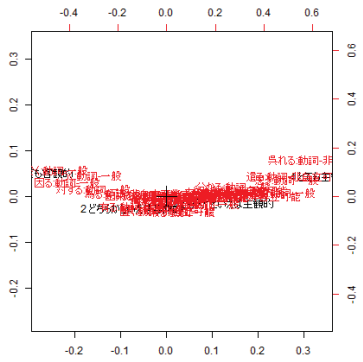
基本的に第 1 主成分（横軸）方向に指標の尺度順に分布している。『1. 専門家向き』『2. やや専門的な一般向き』間が第 1 主成分方向の差分が少なく，『4. 中高生向き』『5. 小学生・幼児向き』間は第 1 主成分方向の尺度が逆転しているが，第二成分方向に差分があり，識別可能である。第 1 主成分の寄与度は 97.1% である。第 1 主成分方向に見ると「例えば」「最も」が『1. 専門家向き』の語彙，「ずっと」「もう」が『5. 小学生・幼児向き』の語彙であることが確認できる。

佐藤・柏野 (2012) は，今回用いた専門度ではなく obi/B9 難易度 (佐藤 (2011)) を用い，4 つのテキストを難易度順に並び替える被験者実験結果と obi/B9 との比較を行っている。被験者実験においては，1000 文字程度に揃える，テキストの NDC が同じものにするなどの工夫が見られる。一方，定量的な評価に終始しており，定性的な分析があまりおこなわれていない。

3.2　客観度

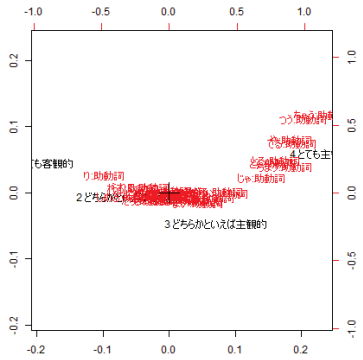
客観度については，保田ほか (2013d) において語りかけ性度との 2 軸の定性的な分析が行われている。以下では動詞・助動詞についての結果を示す。

動詞



基本的に第 1 主成分（横軸）方向に指標の尺度順に分布している。第 1 主成分の寄与度は 97.0% である。「於く」「つく」が『1. とても客観的』の語彙，「遣る」「呉れる」「聞く」が『4. とても主観的』の語彙であることが確認できる。「遣る」「呉れる」については，保田ほか (2013d) の p.151 参考図において言及されている。

助動詞



基本的に第 1 主成分（横軸）方向に指標の尺度順に分布している。第 1 主成分の寄与度は 91.7% である。「り」が『2. どちらかといえば客観的』の語彙，「ちゃう」「てる」「ちまう」「てく」が『4. とても主観的』の語彙であることが確認できる。一方，保田ほか (2013d) では，『4. とても主観的』群に特徴的な語として「たい」をあげているが，本分析では『3. どちらかといえば主観的』あたりに位置している。作業者は「たい」を直接主観度の高い表現とみなしているわけではなく，主観度の高い文書に頻出する「個人的な考え・感情を述べる表現」に頻出する表現としての助動詞「たい」あげていることから，「個人的な考え・感情を述べる表現」でない「たい」の影響があるのではないかと考える。

保田ほか (2013d) においては，主観度毎に出現頻度の高い表現をいくつかあげている。主観度の高い表現として，「個人的な考え・感情を述べる表現」のほかに「主体の経験を表す表現」や「根拠のない断定」をあげている。客観度の高い表現として，「裏付けがある表現」・「数値データの提示」・「伝聞の使用」をあげている。一旦，これらの表現を想定したうえで，各表現の言語の表層的な特徴を考えるという二段階の処理を経て，以下のような特徴をあげている：

主観度が高い事例の特徴

- 数値表現が少ない
- 主体の経験をあらわす受影受動文 (益岡 (1991))
- 伝聞のモダリティを表出しない複合辞「という」

客観度が高い事例の特徴

- 数値表現が多い
- 降格受動文 (益岡 (1991))
- 伝聞のモダリティを表出する複合辞「という」

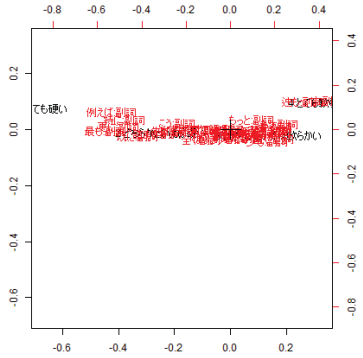
これらの特徴のうち，数値表現については固有表現抽出などの技術や数詞の割合なので判定可能であるが，残りの表現については用法の判定が必要となり何らかの人手によるアノテーションが必要になる。

しかしながらここで重要なのは，作業者のアノテーションにおいては先に述べた表現を探すことを前提として，表層的な特徴をあげている点にある。指標と表現と特徴の 3 つ組を想定してモデル化しなければ，本質的に解けない問題であると考えられる。

3.3　硬軟

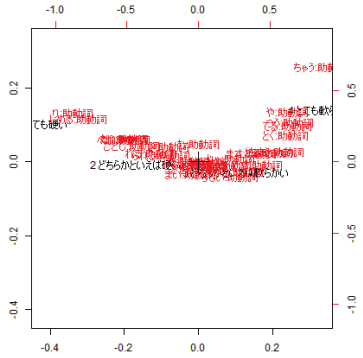
柏野ほか (2012a) は硬軟とくだけ度の 2 軸での定量・定性的な分析を行った。以下では副詞と助動詞と助詞についてとりあげる。

副詞



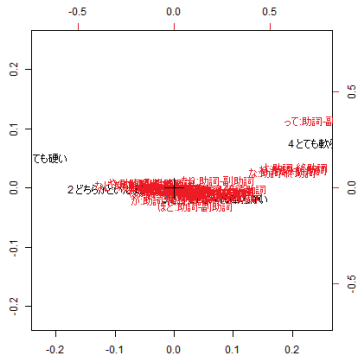
基本的に第 1 主成分（横軸）方向に指標の尺度順に分布している。第 1 主成分の寄与度は 97.7%である。第 1 主成分方向に見ると「例えば」「更に」「最も」が『2. どちらかといえば硬い』の語彙，「迚も（とても）」「ずっと」「一寸」が『4. とても軟らかい』の語彙であることが確認できる。これに対し，柏野・奥村 (2012b) の p.162 表 2 では，「いかに」「より」などの副詞が硬い文書の特徴的表現として挙げられているが，今回処理した頻度上位 30 語にこれらは入っていなかった。柏野・奥村 (2012b) ではさらに，硬い文書では副詞出現頻度が低く，軟らかい文書では副詞のバリエーションが豊富であることについて言及している。

助動詞



基本的に第 1 主成分（横軸）方向に指標の尺度順に分布している。第 1 主成分の寄与度は 92.7%である。第 1 主成分方向に見ると「り」「しめる」が『1. とても硬い』，「ちゃう」「つう」「や」「てる」が『4. とても軟らかい』の語彙であることが確認できる。一方柏野・奥村 (2012b) で言及されている，断定（硬い表現）や「です・ます」（軟らかい表現）などの表現については特別な統計的偏りが見られなかった

助詞



基本的に第 1 主成分（横軸）方向に指標の尺度順に分布している。第 1 主成分の寄与度は 95.0%である。「など : 助詞-副助詞」「や : 助詞-副助詞」が『2. どちらかといえば硬い』の語彙，「ね : 助詞-終助詞」「よ : 助詞-終助詞」「って : 助詞-副助詞」が『4. とても軟らかい』の語彙であることが確認できる。柏野ほか (2012c) は，硬い印象を与える特徴として終助詞がほとんど出現しないことを，軟らかい印象を与える特徴として，特徴的な副助詞（〜か，たり，や，まで）の存在をあげている。前者についてはグラフから読み取れるが，後者についてはグラフから読み取れなかった。

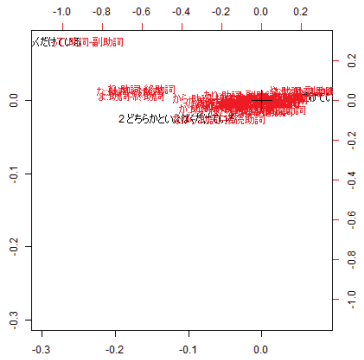
柏野ほか (2012a) は硬度判定の参考情報としていくつかの例をあげている。硬い印象を与える特徴として，「だ・である」調，断定・定義の文，抽象物が主語の受身文，硬い印象を持つ接続詞・副詞の語彙，親密度の低そうな語の頻度，難解な内容，疑問・回答が対応しているなどをあげ，軟らかい印象を与える文の特徴として「です・ます」調，語りかける文末表現，軟らかい印象を持つ接続詞・副詞，平易な語の頻度，平易な内容などをあげている。このうち，今回統計処理で扱っている語彙素のみにより表出するのは一部の文末表現と接続詞・副詞のみである。

また，柏野ほか (2012c) の p.162 表 2 では硬度判定の参考情報がまとめられている。硬い印象を与える特徴として，数詞が多い，文語助動詞などを上記のほかにあげている。

3.4　くだけ度

以下では助詞についての統計処理結果について示す。

助詞



基本的に第 1 主成分（横軸）方向に指標の尺度順に分布しているが，語彙はほぼ中央に分布している。第 1 主成分の寄与度は 96.1% である。「って : 助詞-副助詞」「から : 助詞-接続助詞」「か : 助詞-終助詞」が『1. とてもくだけている』の語彙，「から : 助詞-格助詞」は『2. どちらかといえばくだけている』の語彙，「や : 助詞-副助詞」「の : 助詞-格助詞」「など : 助詞-副助詞」が『3. くだけていない』の語彙であることが確認できる。

柏野ほか (2012a) は，くだけた印象を与える特徴として以下のものをあげている：

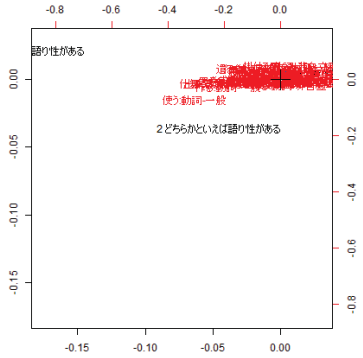
- 体言止めや述語修飾がある
- 一人称が主語の文が多い
- 平易な語に加え，俗語がある
- 卑近な内容や説明である
- オノマトペが多い
- 感覚や感情表現が多い
- 音変化（拗音化・撥音化など）の語がある
- 回答のない，いいっぱなしの疑問文がある

語彙素からは得られにくいものが多く，特に音変化は表層形からは得られるが，語彙素を用いたがために得られない特徴量になっている。

3.5　語りかけ性度

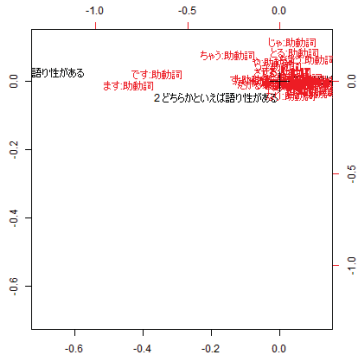
語りかけ性度については，動詞と助動詞を示す。

動詞



基本的に第 1 主成分（横軸）方向に指標の尺度順に分布している。「使う」「作る」が『2. どちらかと言えば語りかけ性がある』付近に分布し，それ以外の語彙は「3. 語りかけ性がない」付近に分布している。第 1 主成分の寄与度は 93.7% である。

助動詞



基本的に第 1 主成分（横軸）方向に指標の尺度順に分布している。「です」「ます」が『2. どちらかと言えば語りかけ性がある』付近に分布し，それ以外の語彙は「3. 語りかけ性がない」付近に分布している。第 1 主成分の寄与度は 99.3% である。

語りかけ性度についても，単純な語彙素による対応分析においてとらえにくいことがうかがえる。語りかけ性については，加藤ほか (2015) が詳細な論考を示している。語彙については，与えられた語りかけ性度に基づいてカイ二乗値を分析し，それぞれの群について有意差のある語を定量的に分析しているほか，実作業者のコメントに基づいて頻度によらない「語りかける」という印象を表出する以下のような特徴的な表現について言及している：

- 昔話に特徴的な表現
- 教示的表現（希望・注意・禁止・勧誘・可能・評価）
- 特定の人称表現（「私たち」「われわれ」）
- 文末表現（「のである」「わけです」「からです」「ものです」，読み手を想定した表現，婉曲表現，読み手の判断を想定した表現）

これらは「語りかける」群にのみ出現率の高い表現ではないか，コーパス全体において出現数が少ないために統計処理によって表出しにくいものであると言及している。

3.6　第 1 主成分の寄与度

表 1 に固定長サンプルの対応分析における第 1 主成分の寄与度を，表 2 に可変長サンプルの対応分析における第 1 主成分の寄与度を示す。可変長サンプルが作業者が見ているサンプルと過不足なく一致するために，統計処理により表出する寄与度も一般的に可変長サンプルの方が高い（表中太字）傾向にある。得られた語彙をみると，寄与度が高い品詞空間において，特別説得力のある語彙素の分布が得られていないこともわかった。

表 1　指標と第 1 主成分の寄与度（固定長サンプル）

	専門度	客観度	硬度	くだけ度	語りかけ性度
連体詞	**91.7%**	96.2%	96.4%	**98.1%**	80.2%
副詞	95.8%	96.2%	96.5%	96.4%	89.7%
助動詞	87.4%	90.3%	92.5%	**88.5%**	99.0%
助詞	91.5%	94.6%	94.4%	**96.4%**	90.7%
動詞	95.1%	96.7%	**97.9%**	**98.8%**	**94.0%**

表 2　指標と第 1 主成分の寄与度（可変長サンプル）

	専門度	客観度	硬度	くだけ度	語りかけ性度
連体詞	**91.7%**	**98.2%**	**97.9%**	98.0%	**91.4%**
副詞	**97.1%**	**97.0%**	**97.7%**	**97.8%**	**92.7%**
助動詞	**88.1%**	**91.7%**	**92.7%**	87.9%	**99.3%**
助詞	**92.9%**	**95.7%**	**95.0%**	96.1%	**90.9%**
動詞	**95.7%**	**97.0%**	97.5%	98.7%	93.7%

4. 考察

専門度・客観度については対応分析によって得られた語彙が先行研究の定性的な分析や作業者の判断基準などと一致する傾向がみられた。一方，硬度・くだけ度については，一部の特徴的な語彙素が得られているが，必ずしも作業者の判断基準としていた参考情報と一致しない部分も見られた。さらに語りかけ性度については語彙素のみによってはとらえられず，アノテーション作業においてより高度な認知的な判断が行われていることが示唆された。

以下，語彙素以外でどのような特徴量を含めるべきかについて検討する。

語彙の印象評定：感覚・感情表現や，語彙そのものの硬軟・くだけ度などが文体判定に用いられている。これらの語彙の分類は単純な形態素解析結果のみからは情報が得られない。単語親密度 (天野・近藤 (1999)) や分類語彙表 (国立国語研究所 (2004)) など様々な語彙分類があるが，文体指標に特化した語彙表の構築が必要である。

オノマトペ・語彙の音変化：オノマトペや語彙の音変化（拗音化・撥音化）など音声言語としての印象が文体に影響を与えている。今回の語彙素による分析では表記ゆれなどを正規化して分析しており，このような音変化が消された分析になっている。発音などの情報を含めることが必要である。

文末表現：特定の文末表現が指標の判定に影響を与えることが示唆されている。文末からの N-gram などを特徴量として含めることにより，指標判定の性能向上がはかれると考える。丸山 (2012a,b) はレジスターごとの文末表現（文末からの N-gram）の傾向を分析調査している。文末表現の構造として，動動詞や終助詞が表出する文法カテゴリ（ヴォイス，アスペクト，肯否，テンス，モダリティ）に注目しているが，実際には 文字 N-gram を展開したうえで，モダリティについての対人（依頼・質問）・対時（禁止・義務・許可）についての出現率の分析を

行っている。しかしながら，BCCWJ のコアデータにはこれらの多様な文法カテゴリの人手によるアノテーションが進められており (松吉ほか (2014), 4.1 節)，これらを用いて分析することが考えられる。

統語的な用法分類：態の用法（受影受動-降格受動）や複合辞が表出するモダリティなどが指標の判定に用いられている。しかしながら，現状のコーパスアノテーションにおいては，このレベルの情報の網羅的な情報付与がされていない。

文の談話的機能：1 文の談話的機能や複数文の談話的機能の推移が文体に影響を与えている。文の談話機能としては，希望・注意・禁止・勧誘・可能・評価・断定・定義・疑問・回答などがあげられている。一部については統語レベルのアノテーション，もしくは言語処理で用いられる応用よりのアノテーションから情報を得られるが，文体そのものを評価するための談話的機能の情報が付与されていない。

内容：文章に含まれる内容そのものが文体に影響を与える。これらについては NDC や C コードなどメタデータの情報から得ることが可能であり，先行研究では NDC や C コード別の分析が多く行われている。一部の指標で NDC と C コードごとの指標の分布の偏りが観察されるが，多様な文体が混在する分類もあり，これらのメタデータが適切な内容分類の粒度かどうかについては議論の余地がある。

各特徴量については，現状の言語資源や言語処理技術により実現可能なものもあるが，何らかのアノテーションが必要なものも多くある。一方で，指標がアノテーションされている状況から，指標から特徴的な構造・用法を統計的に抽出する方法も検討する必要がある。

5. おわりに

本稿では，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 10,551 サンプルに対して付与された専門度，客観度，硬度，くだけ度，語りかけ性度の 5 種の分類指標に対して，品詞ごとに語彙素を特徴量とした対応分析を行い，先行研究で言及されている定量的・定性的分析との比較調査を行った。語彙素のみに基づく手法では，作業者によって認知される文体指標を部分的にしかとらえられないことがわかった。

識別的な手法を用いて，特定の指標のみに表出する低頻度の語彙素については特徴的な語彙素に対して分類を付与することは可能であろう。しかしながら，全体に同じ表現が偏在するが，用法の違いにより表出するような指標については，表層に基づく単純な手法ではとらえられない。今後，既存のアノテーションでとらえられる特徴，新たにアノテーションを行うことでとらえられる特徴などを調査する。さらに，隠れ変数を立てた統計モデルを用いることにより，いままでとらえられなかった特徴量がとらえられるのかについて分析していきたい。

謝辞

本研究の一部は国語研「超大規模コーパス構築プロジェクト」によるものです。

参考文献

天野成昭・近藤公久 (1999).『NTT データベースシリーズ日本語の語彙特性』　三省堂　1 巻.

柏野和佳子・立花幸子・保田祥・丸山岳彦・奥村学・佐藤理史・徳永健伸・大塚裕子・佐渡島紗織 (2012a).「テキストの硬さと柔らかさの考察–『現代日本語書き言葉均衡コーパス』の収録書籍を対象に–」　第 1 回コーパス日本語学ワークショップ, pp. 131–138.

柏野和佳子・奥村学 (2012b).「書籍テキストへの分類指標人手付与の試み–『現代日本語書き言葉均衡コーパス』の収録書籍を対象に–」　言語処理学会第 18 回年次大会, pp. 1260–1263.
柏野和佳子・立花幸子・保田祥・飯田龍・丸山岳彦・奥村学・佐藤理史・徳永健伸・大塚裕子・佐渡島紗織・椿本弥生・沼田寛 (2012c).「書籍テキストへの文体情報付与の試み–『現代日本語書き言葉均衡コーパス』の収録書籍を対象に–」　第 1 回コーパス日本語学ワークショップ, pp. 155–164.
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, pp. 43–53.
加藤祥・柏野和佳子・立花幸子・丸山岳彦 (2015).「語りかける書きことばの表現」　国立国語研究所論集 (印刷中), 8.
国立国語研究所 (2004).『分類語彙表増補改訂版』　大日本図書.
丸山岳彦 (2012a).「『現代日本語書き言葉均衡コーパス』を用いた文末表現のバリエーション」　言語処理学会第 18 回年次大会発表論文集, pp. 591–594.
丸山岳彦 (2012b).「『現代日本語書き言葉均衡コーパス』を用いた文末表現のバリエーション (2)」　第 2 回コーパス日本語学ワークショップ, pp. 207–214.
益岡隆志 (1991).『受動表現と主観性（『日本語のヴォイスと他動性』）』　pp. 105–121.　くろしお出版.
松吉俊・浅原正幸・飯田龍・森田敏生 (2014).「拡張 CaboCha フォーマットの仕様拡張」　第 5 回コーパス日本語学ワークショップ, pp. 223–232.
佐藤理史 (2011).「均衡コーパスを規範とするテキスト難易度測定」　情報処理学会論文誌, 52:4, pp. 1777–1789.
佐藤理史・柏野和佳子 (2012).「テキストの難易度に対する人間の判断と機械の判断」　第 1 回コーパス日本語学ワークショップ, pp. 195–202.
保田祥・柏野和佳子・立花幸子・丸山岳彦 (2012a).「「語りかけ性」を有すると判断される書きことばの表現」　第 2 回コーパス日本語学ワークショップ, pp. 43–50.
保田祥・柏野和佳子・立花幸子・丸山岳彦 (2012b).「「語り性」を有する書き言葉の典型例の分析」　第 1 回コーパス日本語学ワークショップ, pp. 139–146.
保田祥・柏野和佳子・立花幸子 (2012c).「総体として印象を与える表現 :「語りかけ性」を有すると判断する根拠」　人工知能学会第 41 回ことば工学研究会.
保田祥・柏野和佳子・立花幸子・丸山岳彦 (2013a).「「ベテランは足を保護する」が語りかけるとき」　第 4 回コーパス日本語学ワークショップ, pp. 345–354.
保田祥・柏野和佳子・立花幸子・丸山岳彦 (2013b).「アノテーターコメントを用いた「語りかけ性」分析の試み–頻度情報から捉え難いテキスト性質の解明に向けて–」　言語処理学会年次大会発表論文集.
保田祥・柏野和佳子・立花幸子・丸山岳彦 (2013c).「語りかけると判断される文体–大規模コーパスを用いた特徴的表現の分析–」　日本文体論学会第 104 回大会.
保田祥・柏野和佳子・立花幸子・丸山岳彦 (2013d).「書きことばにおける「語りかけ」は何のため用いられるのか」　第 3 回コーパス日本語学ワークショップ, pp. 143–152.