

## 文体指標と語彙系列の対応分析

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 n-gram/p-mer

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

### 要旨

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。浅原ほか (2014) では, この分類指標に対して語彙素を特徴量とした制約付き主成分分析を行い, 各指標と特徴的な語彙分布の対応を品詞ごとに定量的に評価した。本研究では語彙素を語彙の系列 (n-gram, p-mer) に拡張し対応分析を行う。

### 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 (2014))。また, 浅原ほか (2014) では, この分類指標に対して語彙素を特徴量とした制約付き主成分分析を行い, 各指標と特徴的な語彙分布との対応を品詞ごとに定量的に評価した。

本研究ではこの手法を拡張し, 各指標と語彙系列 (語彙素の連続・非連続列) との対応分析を行う。具体的には語彙系列から高頻度の n-gram・p-mer を特徴量として, 文書-特徴量行列を構成し, 制約付き主成分分析を行った。分析結果について報告する。

### 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

表1 文体指標と日本十進分類法 (NDC) の対応

先行研究	文体指標	高い		低い	
		NDC	タイトル例	NDC	タイトル例
柏野・奥村 (2012b)	専門度	1	『大衆社会のゆくえ』『日常生活の認知行動』	9	全般
		3	『注釈会社法』『刑事司法を考える』『アジアの成長と金融』	無	全般
柏野・奥村 (2012b)	硬度	1	『新・岩波講座哲学』『親鸞の思想構造序説』『日本近代思想大系』	7	『競馬探偵の憂鬱な月曜日』『松井秀喜メジャー物語』『芸能界デビュー』『二代目さん』
		2	『西洋中世都市の自由と自治』『日本近代社会成立期の民衆運動』『岩波講座近代日本と植民地』	9	『吉永さん家のカーゴイル』『耽美小説の書き方』『アンネの日記』『小説道場』
		3	『現代国際法』『フランスの社会保障』『情報ネットワーク社会の展開』		
柏野ほか (2012a)	くだけ度	7	『ほんじょの眼鏡日和。』『清水ミチコの「これ誰っ!？」』『関根勲のやるだけマルもうけ!!』	4	『日本の昆虫』『フォッサマグナ』『生体物質とエネルギー』
		9	『ヒミツの転校生』『少しだけラブストーリー』『いつでもこの世は大霊界!』	8	『暗号の数理』『天学で教える小論文の書き方』『語源をさぐる』
加藤ほか (2014)	語りかけ性度	15	『いま、四十代を生きる女へ』『サラリーマンなんか今すぐやめなさい』『敬語マニュアル』	ハウツー本でない本全般	
		49	『消化器ガン克服マニュアル』『読むだけで絶対やめられる禁煙セラピー』『妊娠レッスン』		
		59	『わかりやすいイタリア料理』『お産と育児のスーパーアドバイス』		
		8	『はがきの書き方』『人に好かれることばレッスン』『講師・講演を頼まれたら読む本』		
柏野・奥村 (2012b) 保田ほか (2013d)	客観度	1	『発達心理学入門』『認知とパフォーマンス』『宗教改革』『岩波講座東洋思想』	2	『インドでわしも考えた』『泣いたら負けや』『ほいほい旅団ベトナムへ行く』
		4	『超分子化学への展開』『生物科学入門』『環境化学』『ガン全種類別・最新治療法』	9	『日はまた熱血ボンちゃん』『マンボウ交友録』『パパは塾長さん』『男はオイ!女はハイ…』

NDCは日本十進分類法(NDC分類)の上位桁を表す。無はサンプルにNDC分類が付与されていないもの。

#### 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代女性作業員延べ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。

表1に先行研究で言及されている文体指標と日本十進分類法(NDC)の対応をタイトルとともに示す。

## 2.2 特徴量の取得

本研究では語彙系列の  $n$ -gram (連続部分列) と  $p$ -mer (非連続部分列) を特徴量とする。 $n$ -gram とは系列に対する長さ  $n$  の連続部分列 (substring) のことをいい、 $p$ -mer とは系列に対する長さ  $p$  の部分列 (subsequence) のことをいう。

例えば“ABCDE”という系列に対して、3-gram は“ABC”, “BCD”, “CDE”の3種類あり、3-mer は“ABC”, “AB/D”, “AB/E”, “A/CD”, “A/C/E”, “A/DE”, “BCD”, “BC/E”, “B/DE”, “CDE”の10種類あり、それぞれ頻度は1である。 $p$ -mer の“/”は、そこにギャップがあることを意味している。

系列特徴量の枚挙には、系列パターンマイニングアルゴリズム `prefixspan`<sup>(1)</sup>(Pei et al. (2001))を用いる。系列特徴量の枚挙の詳細な仕様については以下の通り：

- 系列中、同じパターンが複数回出現する場合は全ての組合せを枚挙する。“ABABAB”の中に“AB”は3回、“A/B”は3回出現したと数える。
- ギャップ長については制限を課さない。
- ギャップの挿入箇所が違う  $p$ -mer は別のものとして扱う。
- 枚挙は頻度 10 以上のものを全てを展開した。

なお、文体指標付与時には台詞をのぞいた地の文のみを対象としているが、特徴量の展開は台詞・地の文の識別を行わず、可変長サンプル全体から行った。

## 2.3 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たかの尺度となる。同様に第1主成分から第 $\alpha$ 主成分までの寄与率の和を第 $\alpha$ 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たかの尺度となる。

実際の計算には R の MASS パッケージを用いた。`prefixspan`により枚挙した特徴量を R のデータフレームのカタチで表現したあとは、R のプログラムとしては数行のものである。この数行を Latent Dirichlet Allocation (Blei et al. (2003)) のパッケージに置き換えることにより、語彙系列-文書間のトピックモデルを作成することも可能である。

(1) 実装は <http://prefixspan-rel.sourceforge.jp/> を用いた。

表 2 専門度と語彙系列の対応分析結果 (第 1 主成分の上位 5 件・下位 5 件)

2gram 下位	3gram 下位	4gram 下位	5gram 下位
使用-者	適用-為る-れる	べし-だ-有る-と	れる-ない-ば-成る-ない
所有-権	労働-者-の	た-場合-に-は	次-の-様-だ-述べる
債権-者	規定-為る-れる	に-於く-て-は	化-為る-れる-て-いる
結合-為る	有する-て-居る	示す-れる-て-居る	構成-為る-れる-て-居る
公共-団体	た-場合-に	れる-ない-ば-成る	的-だ-物-だ-ある
2gram 上位	3gram 上位	4gram 上位	5gram 上位
御-化け	が-言う-ます	が-言う-ます-た	成る-て-居る-ます-た
姉-ちゃん	御-姉-ちゃん	御-婆-ちゃん-は	そう-だ-顔-を-為る
御-婆	婆-ちゃん-は	御-父-さん-は	て-仕舞う-た-の-です
兄-ちゃん	言う-ます-た	て-行く-ます-た	何時-も-の-様-だ
婆-ちゃん	御-婆-ちゃん	の-御-父-さん	どう-為る-た-の-だ
4mer 下位	5mer 下位	6mer 下位	7mer 下位
的的為る為る	的にて為る為る	はの的のの為る	のの為る為るの為る為る
於くの為に為る	為るの的為る為る	の的のの為に為る	のの為るとの為る為る
的に的為る	にて的為る為る	はの的の為に為る	の為に為るの為る為る
於くはの為に為る	に的の為に為る	的のののを為る	の為る為るのの為る為る
的的の為に為る	に的に為る為る	のの的に為る	の為るの為るの為る為る
4mer 上位	5mer 上位	6mer 上位	7mer 上位
御さんてた	はてをます-た	はのをてます-た	はのののをてます
さん-はをた	のて-居る-ます-た	はをてをます-た	ののののをてます
さん-はてた	はててます-た	をてをてます-た	のにてのをてます
さん-はをて	をててます-た	はてをてます-た	はのののをてます
はをます-た	はをてます-た	はをててます-た	のをてのをてます

### 3. 結果と考察

#### 3.1 結果

結果のテーブルは大部なため、<http://goo.gl/4aFQTj> に公開する。参考のため表 2 に専門度と語彙系列の対応分析結果のうち第 1 主成分の下位 5 件・上位 5 件を示す。

#### 3.2 考察

これらの文体指標の分析例には柏野・奥村 (2012b) があり、NDC 別の特徴をまとめている。先行研究に見られるサンプル群の NDC 分布の傾向から、どのような書籍が特徴的と考えられる NDC 群に含まれているのかを調べ、本手法で取得された表現が実際どのように現れているのか、特に 3gram, 4gram, 5gram, 4mer について例示する。

硬度とくだけ度については、柏野ほか (2012a) が、目視によって得られた文末・主語 (一人称)・語彙 (親密度が低そう・平易) などの特徴を列挙している。文末表現や主語については、本手法で得られた特徴と重なる部分がある。語りかけ性度については加藤ほか (2014) が特徴的表現を調査している。また、保田ほか (2013d) は、語りかけ性度との関連において、客観度の高低が、データ・伝聞などの表現に見る根拠の有無によって判断されているとの観点で客観度に関わる特徴的表現例を示した。先行研究で示された表現との異同を確かめ、本手法で新しく捉えられた特徴と捉えられなかった特徴を確認する。

##### 3.2.1 専門度

最初に専門度で得られた語彙系列特徴量の第 1 主成分の偏りについて示す。

- 3gram

「条+数詞+項」「適用される」「規定される」「対象とする」「裁判所の」「有している」「於いては(も)」「主義的だ」「合理的だ」「目的とする」⇔「お姉ちゃん」「おばあさん」

「たのかしら」「のだぞ」「たわね」「のかい」「なんだか」「私だって」「どうかする」

- 4gram

「た場合には」「原則として」「たものとする」「であるとする」「べきであると」「示されている」「前提として」「明らかにされる」「することによる」「なされている」⇔「が言いました」「て行きました」「やって来ます」「たのでした」「ちゃったのだ」「って言うのは」「どうしたの」「もしかして」「あっと言う間」

- 5gram

「れなければならない」「するのではない」「構成されている」「行われていた」「次のように述べる」「事を示している」「を意味している」「されることになる」「ということである」「であるという」⇔「してしました」「てしまったのです」「ありませんでした」「のでないかな」「いつの間にか」「なんとかして」

- 4mer

「的～的～する～する」「的～の～的～の」⇔「さんは～を～た」「さんが～て～た」「その～に～ました」

このうち 4mer の「的～的～する～する」「的～の～的～の」のような一文中に「的」が頻出するパターンは過去の研究（頻出語彙や人手による検討）で捉えられなかった特徴である。

例 (LBi3\_00033 『現代法社会学入門』 321)<sup>(2)</sup>

前述のように、パレート最適という価値判断基準とコースの定理によって、法的ルールや制度について効率性の観点からの理論的分析や実証的研究、さらには規範的な問題提起や提言をすることが可能となる。

以下の例では特徴的な語彙系列を含む専門度の高い例を示す。

例 (LBj1\_00010 『日常生活の認知行動』 141)

議論の余地があるのは、もしひとつの方法だけがすべての問題に自動的に 適用されるなら、さまざまな比率の比較の意味をきちんと保ったままているのはもっとやさしいだろうということである。しかしながら、人々は、このようなやり方で自分たちの日常生活を単純化してしまうことはない。

以下では特徴的な語彙系列を含まないが、専門度が低い例を示す。この例では語彙系列ではなく表記から専門度が低いと判定されたと考えられる。

例：小学生・幼児向き (LBcn\_00024 『宇宙人はほんとにいるか?』 NDC 分類無)

この天の川銀河にどれくらいの数の惑星系があるのかはまだわかりませんが、これまでの観測結果からはずいぶんあるだろうという予想がたっています。そのなかには、地球に似た、生命をやどしている星もあるにちがいありません。

以下では特徴的な語彙系列を含まないが、専門度が高い例を示す。語彙が分野的に特化している場合や、表現の共起関係が一般的でない場合は特徴が表れにくい。

例：(LB10\_00010 『最終講義』 041)

これを政治史でみますと、中国の国民革命軍が北伐を完成したのは一九二八年、いまま

<sup>(2)</sup> 括弧内は、サンプル ID・『タイトル』・NDC 分類。以下同様。



で分裂していた中国が北伐の完成によって統一され、軍閥、とにかく国民党が南京に都を置いて、中央政府ができてきました。そのもとで中央研究院歴史語言研究所というような研究機関が設立され、そこで安陽の発掘ができ、またいまままで政治闘争にあけくれていた大学でありましたが、古典の研究などに本腰を入れてやりだした。

### 3.2.2 硬度

硬度について第1主成分の偏りのある語彙系列を確認する。

まず、2gram でレジスタ特徴の見られる名詞的表現が得やすい：

「条約の」「権の」「規定に」「著作権」「債権者」「使用者」「公共団体」「損害賠償」

さらに、3~5gram および 4mer では名詞的表現だけでなくレジスタ特徴のあるコロケーションが得られている：

- 3gram

「条+数詞+項」「労働者の」「重要な役割」「第一次」「委員会が」「裁判所の」「市町村」「我が国の」「総合的+だ(助動詞)」「普遍的+だ(助動詞)」「全国的+だ(助動詞)」「象徴的+だ(助動詞)」「有している」「提出される」「問題となる」「原則とする」「ように述べる」「化すること」「べきである」「するものと」「大きな影響を」「明らかである」「既に述べた」「結果である」⇔「なっちゃった」「って言われる」「こうやって」「思ってた」「だよ」「僕らの」「たのよ」「お姉ちゃん」「お兄ちゃん」

- 4gram

「するものである」「重要な役割を」「すべきである」「あるとして」「と述べている」「役割をはたして」「のであるから」「これに対して」「都道府県」「第二次大戦」⇔「が好きなの」「してたの」「かなと思う」「なのですよ」「でも私は」「ありがとうございました」「そんな事を言う」

- 5gram

「次のように述べる」「存在していた」「よく知られている」「このようにして」「なったのである」「おかなければならない」「第二次世界大戦」⇔「かもしれないけれど」「のではありません」「いませんでした」「聞いたことがある」

- 4mer

「於いて~する~て」「於いて~の~が」⇔「私は~ました」「僕は~て~た」

このうち 4mer の「於いて~する~て」「於いて~の~が」のようなパターンは過去の研究(頻出語彙や人手による検討)で捉えられなかった特徴である。

例(LBp1.00005『連続性の哲学』133)

いかなる演繹に おいても その実験に おいて 生じる全過程は、厳密にこの三つのもの、すなわち、総括、複製、消去である。それ以外には結論の観察が残るだけである。ただし、これらの三つの実験の要素は、あらゆる演繹に おいて 生じるわけではない。

以下では特徴的な語彙系列を含む硬度が高い例を示す。

例(LBi2.00076『弥生の王国』210)

復元にあたって問題になるのは尺度のことである。わが国古代には人体尺で建物を測っていたということを前提にして、このたびの復元も人体尺によることにした。そのこと

について説明して おこななければならない。

### 3.2.3 くだけ度

くだけ度の第1主成分の偏りのある語彙系列を確認する。

- 2gram

「しちゃう」「けれどさ」「てさ」「ちゃって」「かよ」「ちゃん+助詞」「こと言う」「訳ない」「よな」「だぜ」「ちまった」「てるって」⇔「意思決定」「連合国」「出土する」「軍事的」「東アジア」「労働組合」「安全保障」「削除する」「分布する」「承認する」「経済学」

- 3gram

「お姉ちゃん」「つうことは」「てるのよ」「言ってた」「やってるの」「なっちゃった」「だからね」「そうだよ」「俺には」「だったな」「ですよ」「馬鹿にする」⇔「事が明らか」「例である」「総合的だ」「進められて」「形成される」「中小企業」「実質的な」「記している」「理解される」「我が国で」

- 4gram

「のだよね」「なのだけれど」「とと思ってた」「どうでも良い」「気にしない」「のだなど」「してくれない」「だとすると」⇔「示されている」「対象として」「指摘している」「とされた」「にあたっては」「原則として」「認められている」全体として」「具体的には」

- 5gram

「かもしれないけれど」「でないかと思う」「な気がした」「見たことがない」「なっていました」「それにしても」「身につけている」⇔「と考えられている」「的なものである」「事を示している」「よく知られている」「行われていた」「られたのである」「ようなものである」

- 4mer

「僕は～て～た」「そんな～を～た」「ような気～する」「目を～て～た」「どうして～の」「だから～のだ」⇔「於いて～に～する」「的～を～的～する」「このよう～の～する」

以下では特徴的な語彙系列を含む、くだけ度が高い例を示す：

例 (LBn7\_00044 『ほんじよの虫干。』 770)

堀辰雄 (『風立ちぬ』) と並んで、当時“四季派”と呼ばれた詩人達の中の代表的な人物。建築デザイナーでもあったんだよ。カッコいい！初めてその詩に触れた時には、そのあまりの美しさに感激して、同時にそのあまりの儂さに (大丈夫?) って驚いたものでした。なよなよとしててねえ、デリケートでナイーブで神経質で、こりゃ早く死ぬのも無理ない よなあ、って思わせるの。

以下では特徴的な語彙系列を含む、くだけ度が低い例を示す：

例 (LBi4\_00028 『岩波講座現代の物理学』 420)

われわれは、なにゆえに、自由な Dirac 場が Fermi 統計をみだし、自由な Klein-Gordon 場が Bose 統計に従うかをみてきた。相対論的な自由場はこれ以外にもさまざまなスピンを記述するものが知られている。これらを逐一述べるのは本書の目的ではない。ただスピンと統計の関係として次の定理 (W.Pauli, 1940) は よく知られている。





これは先程分類しました鏡が、今度は各地でどういうふうに散らばって古墳で出てくるか という ことを示したものであります。たとえば一番古い鏡 という のは、その表で言いますと、山陽、畿内、三重、東海、関東とゴシックで書いておりますように、多い です ね。パーセントで言うと、畿内で三八%、九州にもあります。

### 3.2.5 客観度

#### ● 3gram

「条一項」「をクリックする」「設定される」「原則とする」「消費者の」「定められて」「困難である」「確認される」「場合には」「理由とする」「これによる」⇔「だったな」「そ  
うだよ」「のですって」「私のこと」「生まれて初めて」「好きになる」「なんだか」「てい  
るのだ」「そうかも」「僕には」「私はもう」

#### ● 4gram

「用いられている」「ことを示して」「とも呼ばれる」「たことがわかる」「重要な役割を」  
「数詞（千九百～）」「例として」「発見された」「を明らかにする」⇔「のだよね」「でな  
いかな」「言ってるのだ」「ちゃったのだ」「でも私は」「が好きなの」「かなと思う」「な  
んとかなる」「気がした」「私にとって」「のだそうだ」「おぼえていない」

#### ● 5gram

「と考えられている」「事を示している」「することができます」「となっています」「れな  
ければならない」「とされてきた」「として知られる」⇔「のでないかな」「と思ったの  
だ」「とは思わなかった」「いつものようだ」「たのだと思う」「聞いたことがある」「たこ  
とがあります」

#### ● 4mer

「因る～が～する～する」「因る～の～の～れる」⇔「僕は～」「私は～」「私が～」「私の  
～」

以下では特徴的な語彙系列を含む、客観度が高い例を示す：

例 (LBa3\_00047 『地方財政の国際比較』 349)

しかもジョーンズらに よれば, 左右を問わず地域内外 の 政治的党派的集団 の 動き が 地方団体政策決定 の 焦点となり, 伝統的な委員会制度 の 機能が弱まりつつあるなど (64), 地方団体の運営にもこれら政治状況の影響が及んでいる。

例 (LBo4\_00040 『超分子化学への展開』 431)

細胞外から の シグナル分子 が 膜受容体に結合 する と, Gタンパク質の働きに よって, 細胞内部 の 酵素 が 活性化 さ れる. 人工脂質膜上に人工受容体と酵素を固定化 する と, シグナル化合物に よ っ て酵素活性 が 増加 する 人工システムを構築できる。

## 4. 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、品詞ごとに語彙系列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。

## 謝辞

国語研基幹型共同研究プロジェクト「コーパスアノテーションの基礎研究」および国語研「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

- 浅原正幸・加藤祥・立花幸子・柏野和佳子 (2014). 「文体指標と語彙の対応分析」 第6回コーパス日本語学ワークショップ, pp. 11–20.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003). “Latent Dirichlet Allocation.” *Journal of Machine Learning Research*, 3, pp. 993–1022.
- 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). 「テキストの硬さと軟らかさの考察—『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 (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 (2014). 「語りかける書きことばの表現」 国立国語研究所論集, 8, pp. 85–108.
- Pei, Jian, Jiawei Han, Behzad Mortazavi-Asi, Helen Pinto, Qiming Chen, Umeshwar Dayal, and Mei-Chun Hsu (2001). “PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth.” *Proceedings of the 17th International Conference on Data Engineering*, pp. 215–224.
- 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」を有する書きことばの典型例の分析」 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・立花幸子・柏野和佳子・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—」 言語処理学会第19回年次大会発表論文集, pp. 358–361.
- 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」は何のために用いられるのか」 第3回コーパス日本語学ワークショップ, pp. 143–152.