様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成　24 年　6 月 11 日現在

機関番号：８２６３６
研究種目：若手研究（B）
研究期間：2009～2011
課題番号：　21720166
研究課題名（和文）　日本語「書き言葉らしさ・話し言葉らしさ」測定法の設計

研究課題名（英文）　Establishing measurements for written and spoken likeness of texts

研究代表者
佐野　大樹　（SANO　MOTOKI）
情報通信研究機構ユニバーサルコミュニケーション研究所・情報分析研究室・専攻研究員
研究者番号：60455425

研究成果の概要（和文）：本研究は、日本語のテクストの「書き言葉らしさ・話し言葉らしさ」の測定法を設計し、これを評価、応用するものである。「書き言葉らしさ・話し言葉らしさ」を、単に文字媒体・音声媒体に固有な特徴と捉えず、選択体系機能言語理論に基づき、①テクストにおける情報の密度、②テクストで扱われる事象の具体性・一般性、③テクストに表される感情・意見・態度の主観性・客観性の 3 つの観点から捉えて計測法を構築し、これを『現代日本語書き言葉均衡コーパス』を用いて評価した。さらに、構築した計測法を用いて『日本語アプレイザル評価表現辞書』などの言語資源を編纂し、これを公開した。

研究成果の概要（英文）：This study explores the measurements for understanding written and spoken likeness of Japanese texts. The study employed systemic functional theory as its framework, and examined the written and spoke likeness of texts from three perspectives, which are i) lexical density of texts, ii) degree of (de)contextualisation and iii) tendencies of the usages of attitudinal lexis. Furthermore, this study evaluated the constructed measurements using the Balanced Corpus of Contemporary Written Japanese and also applied the measurements in order to compile linguistic resources such as 'Japanese Dictionary of Appraisal'.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2009 年度	900,000	270,000	1,170,000
2010 年度	500,000	150,000	650,000
2011 年度	700,000	210,000	910,000
年度			
年度			
総　計	2,100,000	630,000	2,730,000

研究分野：言語学
科研費の分科・細目：言語学・日本語学
キーワード：書き言葉・話し言葉・語彙密度・コーパス・アプレイザル・科学リテラシー・辞書・評価

１．研究開始当初の背景
英語のコーパス研究においては、「書き言葉らしさ・話し言葉らしさ」を計測する方法が機能言語理論、語用論、談話分析などの枠組みで提示されている。例えば、Systemic Functional theory（選択体系機能言語理論）の枠組みでは、Lexical Density（語彙密度）計測法が提案され、この計測法は言語学、教育学だけでなく、医療分野でも患者にわかりやすいリーフレットの作成過程において活用されている。
　一方、日本語研究においては、「書き言葉・話し言葉」の特徴について検討した研究は多数あるものの、大規模コーパスに対して「書き言葉らしさ・話し言葉らしさ」を計測する手法は確立されていない。この背景には、日本語の大規模コーパスが近年まで整備されていなかったこと、また、「書き言葉・話し言葉」と「書き言葉らしさ・話し言葉らしさ」が理論的に区別されることが少ないことなどがあった。

２．研究の目的
本研究では、日本語の「書き言葉らしさ・話し言葉らしさ」の計測方法を設計するため、以下３つを行った。

(1)英語研究などで用いられている既存の計測方法を用いた「書き言葉らしさ・話し言葉らしさ」の測定
(2)既存の計測方法の評価及び問題点の整理（言語学的分析・統計的手法を用いた検証）
(3)有効な計測方法の設計、及び、論文発表やWebによる発信

３．研究の方法
(1)コーパスについて
本研究で用いたコーパスは、主に以下の２つである。

①『現代日本語書き言葉均衡コーパス』（BCCWJ）
BCCWJは、国立国語研究所を中心に開発された日本語で初の大規模均衡コーパスであり、書籍、新聞、雑誌、ブログ、韻文、白書等、様々な媒体のテクストによって編纂されている。約１億語の規模をもち、基本的には文字媒体のテクストが収録されている。物語文、報告、説明文、主張文など多種多様なジャンルを含む。
②「健康と病いの語りデータベース」
「健康と病いの語りデータベース」は厚生労働科学研究費補助金がん臨床研究事業「がん患者の意向による治療方法等の選択を可能とする支援体制整備を目的とした、がん体験をめぐる「患者の語り」のデータベース」プロジェクトを中心に構築されたもので、maximum variation sampling法に基づいて選出された乳がん患者、前立腺がん患者のインタビューが収録されている。乳がん患者、前立腺がん患者のインタビュー合わせて、約160万語が収録されている。

BCCWJは、日本十進分類法などを基準とした層別無作為サンプリングによって現代日本語書き言葉の縮図となるように設計されており、設計した測定法の汎用性を評価する上で有効であると考えた。また、「健康と病いの語りデータベース」には、がんの検診・治療等に関する専門的な内容から、患者の日常生活に関する内容まで含まれており、設計した計測法の具体的応用について検討する上で有効だと考えた。

(2)「書き言葉らしさ・話し言葉らしさ」計測の観点
本研究では、「書き言葉らしさ・話し言葉らしさ」を、単に、文字媒体・音声媒体に固有の特徴と捉えずに、選択体系機能言語理論を基底として以下の３つの観点から計測した。

①テクストにおける情報の密度
②テクストで扱われる事象の具体性・一般性
③テクストに表される感情・意見・態度の主観性・客観性

これら３つの観点を用いたのは、これらの観点と、同義語（「激しくなる」と「激化する」など）の選択や「書き言葉らしさ・話し言葉らしさ」の印象評定データとに、連関があることが確認されたためである。

(3)「書き言葉らしさ・話し言葉らしさ」の計測法の構築
選択体系機能言語理論で英語のテクスト分析に用いられている手法を参考に、日本語「書き言葉らしさ・話し言葉らしさ」の計測法を構築した。具体的には「テクストにおける情報の密度」については語彙密度、「テクストで扱われる事象の具体性・一般性」については修辞ユニット分析、「テクストに表される感情・意見・態度の主観性・客観性」についてはアプレイザル分析を基底として計測法を構築することにした。

４．研究成果
(1) テクストにおける情報の密度の計測
M.A.K.Hallidayによって提案されたlexical density 計測法について、日本語にも適用可能かBCCWJの書籍データを用いて検証した後、日本語で自動計測ができるようにした。語彙密度は、当該のテクストにおける内容語

の総数を節（文法単位のひとつ）の総数で除算することによって求められる。この手法を活用し、BCCWJ に含まれるテクストの語彙密度を計測し、特定の語が使用されるテクストの語彙密度平均値を求めることで、当該の語の「かたさ」「やわらかさ」を客観的データに基づき示すことができるようになった。例えば、「じれったい（焦れったい）」「いらだつ（苛立つ）」「しょうりょ（焦慮）」などの語彙密度平均値は、順に「3.19」「3.52」「4.02」となる。これによって、辞書に同義語の位相情報を付与したり、かたい難解な表現をやわらかい表現に言い換えたりすることができるようになる。応用の一例として、語彙密度平均値を、類似した価値基準を示す表現が言語使用においてはどのように異なるのかを説明するための指標の一つとして『日本語アプレイザル評価表現辞書』に付与し、公開した。なお、このような試みは、英語の評価表現辞書でも行われておらず、本研究が初めて達成した成果の一つである。

(2) テクストで扱われる事象の具体性・一般性の計測
C.Cloran によって提案された Rhetorical Unit Analysis（修辞ユニット分析）について、日本語にも適用できるよう、BCCWJ ブログデータと 3 名の評定者による「書き言葉らしさ・話し言葉らしさ」の印象評定の結果を踏まえて、再構築した。修辞ユニット分析では、テクストをメッセージ（意味単位。基本的には「節」と類似。）に分割し、

①各メッセージの中心となる要素と話し手・書き手との空間的距離
②メッセージで表される出来事が起きた時間とメッセージが発信された時間との相対的な距離

から、メッセージの談話機能を分類する。例えば、「私が受けた○○治療は、ちょっと痛かったです」というメッセージであれば、中心となる要素が「私が受けた○○治療」で書き手・話し手が実際に経験した行為（状況内要素）で、メッセージが発信された時間よりも過去の出来事について示すので談話機能は、「回想」と分類される。これに対して、「○○治療は痛みを伴う」というメッセージであれば、中心となる要素が「○○治療」で特定の患者が経験した○○治療でなくあるクラス（この場合、○○治療のインスタンスの集合）に属するもの全てを対象とし（定言要素）、メッセージが発信された時間は特定の時間軸に制約されない（習慣的・恒久）ため、談話機能は、「一般化」と分類される。このような方法で、テクストにおいて使用されている談話機能を分析し、その割合などからテクストで扱われている事象の具体性・一般性について計測することができる。応用の一例としては、専門日本語教育分野などのエッセイライティングなどでの活用が期待される。日常的な経験などの記述が多く専門性が確立できていないエッセイを、談話機能を変更することによって修正し、専門性を確立するなど作文指導で利用できる。また、医療分野や原発問題などで顕在化した、専門性が高く一般には分かりにくい情報を、どのように具体化し分かりやすくできるのか検討する上でも、活用できる分析手法である。なお、C.Cloran が提案した手法は、基本的に話し言葉の分析法として提案されたものだが、本研究では、書き言葉にも適応できるように再構築してある。

(3) テクストに表される感情・意見・態度の主観性・客観性の計測
J.R. Martin によって提案された Appraisal Analysis （アプレイザル分析）を日本語に適応できるように再構築した。また、再構築した分析法の妥当性について BCCWJ を用いて評価した。再構築したアプレイザル分析では、評価表現を「楽しい」「悲しい」などの内評価（評価者の感情や感情的態度を示すことで評価を表すもの）と「肝要」「非人道的」などの外評価（評価対象の特徴を示すことで評価を表すもの）など、価値基準ごとに評価表現を分類できるようにした。価値基準ごとに評価表現を分類することで、「あの説明は腹立たしい」の「腹立たしい」（内評価の細分類である「情動」に該当）ように主観的に評価を表すもの、「あの説明は抽象的だ」の「抽象的」など「腹立たしい」などに比べて客観的に評価を示すもの（外評価の細分類である「性質」に該当）を区別できる。また、本研究では、分類方法を提案するだけでなく、評価表現約 8000 件を再構築した分析法に従い分類した『日本語アプレイザル評価表現辞書』(JAppraisal 辞書)を構築し、言語資源協会
(http://gsk.or.jp/catalog/GSK2011-C/catalog.html)より公開した。JAppraisal 辞書を用いることで、テクストで使用されている評価表現を認識し、価値基準別（約 200 種類）に分類することが可能である。なお、Appraisal 分析を英語以外の言語に適用、当該の言語特有の分類を構築し、評価表現辞書を公開したのは、本研究が初めてである。

(4)まとめ
「テクストにおける情報の密度の計測」については、日本語における語彙密度の自動計測方法を構築し、辞書構築の際などに、同義語の位相情報として付与できるようにした。「テクストで扱われる事象の具体性・一般

性」については日本語における修辞ユニット分析の適用法を確立し、専門日本語教育などで活用できるようにした。「テクストに表される感情・意見・態度の主観性・客観性」については、日本語における評価表現の分類法を考案し、また、『日本語アプレイザル評価表現辞書』を編纂し、テクストに含まれる評価表現を自動認識、分類する言語資源を公開した。

(5)今後の展望
本研究では、日本語の「書き言葉らしさ・話し言葉らしさ」の計測方法について検討し、英語の手法を再構築することなどにより 3 つの分析手法を設計した。今後は、これらの分析法を用いて、書き言葉らしいテクストを話し言葉らしいテクストに変換するシステムを構築していければと考えている。特に、原発問題などで顕在化した科学リテラシーに関する問題の解決に貢献できるよう、情報の受信者だけでなく発信者に対する教育的視点も含めて検討していく予定である。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　8 件）

① 佐野大樹、アプレイザル理論を基底とした評価表現の分類と辞書の構築、国立国語研究所論集、査読有、３巻、2012、53-83、http://www.ninjal.ac.jp/publication/papers/03/
② 佐野大樹、小磯花絵、現代日本語書き言葉における修辞ユニット分析の適用性の検証−「書き言葉らしさ・話し言葉らしさ」と脱文脈化言語・文脈化言語の関係−、機能言語学研究、査読有、６号、2012 、59-81、http://www.jasfl.jp/journal.html#jjsfl6
③ 佐野大樹、選択体系機能言語理論を基底とする特定目的のための作文指導方法について、専門日本語教育研究、査読無（依頼論文）、12 号、2011、19-26、http://wwwsoc.nii.ac.jp/stje/lang-ja/magazine/backnum_3.html

〔学会発表〕（計　14 件）

① Sano, M. Reconstructing English System of Attitude for the Application to Japanese: An Exploration for the Construction of a Japanese dictionary of appraisal, 38th International Systemic Functional Congress, July 26, 2011, University of Lisbon, Portugal.
② 佐野大樹、修辞ユニットを用いた書き言葉の分析－「書き言葉・話し言葉らしさ」と(脱) 文脈化の関係－、社会言語科学会第25回研究大会、2010年3月13日、慶応義塾大学、東京都

〔その他〕
ホームページ等
① 日本語アプレイザル評価表現辞書、http://gsk.or.jp/catalog/GSK2011-C/catalog.html

６．研究組織
(1)研究代表者

佐野　大樹（SANO MOTOKI）
（独）情報通信研究機構ユニバーサルコミュニケーション研究所・情報分析研究室・専攻研究員
研究者番号：60455425

(2)研究分担者
（　　）

研究者番号：

(3)連携研究者
（　　）

研究者番号：