

平成 22 年 3 月 10 日現在

研究種目：基盤研究(B) 機関番号:13902

研究期間：2007～2009

課題番号：19300280

研究課題名（和文） WWWを用いた日本語コーパスの構築と文章理解支援への応用

研究課題名（英文） The development of WWW corpus for Japanese language education

研究代表者

野崎 浩成 (NOZAKI HIRONARI)

愛知教育大学・教育学部・准教授

研究者番号：80275148

研究成果の概要（和文）：本研究では，WWW 上で流通する日本語テキストを言語資源として活用し，言語構造の分析や意味解析などの計量的な言語処理を行う手法を考案する。そして，外来語を始めとする語彙の用例を収集・分析し，それらを日本語コーパス（コンピュータで処理可能な電子化テキスト）として活用する研究を行う。これにより，外来語などの語彙の使用形態や表現の特徴を明らかにできることから，国語辞典に掲載する新出の外来語の選定や用例集の抽出，難解な IT 用語の概念理解の支援などが実現できるため，言語学のみならず国語教育や情報教育への応用も期待される。

研究成果の概要（英文）：There is a strong demand for the application of Japanese-language text circulating on the WWW as a language resource, to construct a method of quantitative language processing that involves steps such as language structure analysis and meaning parsing, focusing on Japanese in electronic form (particularly text data on the Internet). It is also thought necessary to collect and analysis examples of the usage of vocabulary, including loan words, and research the application of such examples as a Japanese-language corpus (electronic text that can be processed by computers). Since the implementation of such research would enable clarification of the characteristics of usage and expression of vocabulary such as loan words, this research is expected to be applicable not just to linguistics, but also to Japanese-language education and information education in order to implement programs such as the selection of newly-coined loan words that appear in Japanese-language dictionaries and the extraction of collections of usage examples to assist in the comprehension of difficult-to-understand IT language concepts.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2007 年度	5,400,000	1,620,000	7,020,000
2008 年度	4,000,000	1,200,000	5,200,000
2009 年度	3,700,000	1,110,000	4,810,000
年度			
年度			
総計	13,100,000	3,930,000	17,030,000

研究分野：教育工学

科研費の分科・細目：科学教育・教育工学

キーワード：語彙調査，テキストマイニング，日本語教育，テキストデータベース，文章理解，Web，コーパス

### 1. 研究開始当初の背景

一般になじみの薄い「外来語」(アジェンダやサーベイランスなど)を多用することは、高齢者や日本語学習者の文章理解を妨げる大きな要因の1つとなっている。このような情報弱者を支援するために、『「外来語」言い換え案』(国立国語研究所 2003)では、一般に分かりにくい外来語を日本語に言い換える提案をしている。一方、情報技術の急速な伸展とともに、インターネットを始めとする電子テキスト上では、新しいコンピュータ用語(IT用語)が次々と誕生している。このようなIT用語は、英語を語源とする「外来語」が大多数を占めており、コンピュータを専門としない一般読者にとって理解し難い語彙となっている場合が多い。

このように、国立国語研究所が中心となって、「分かりにくい外来語」を「平易な日本語」に言い換える案が出されている一方で、IT用語などWWW上で次々と生まれてくる「新出の外来語」に着目した研究は皆無であった。また、日本語テキストは、国内に限らず全世界規模で、インターネットを介して流通しているにも関わらず、WWW上の日本語そのものを「データ」としてとらえ、それらを言語情報処理の分析対象とする研究が十分に行われてきたとは言えない。

### 2. 研究の目的

本研究の目的は、(1) WWW上で流通する日本語テキストを言語資源として活用する、(2) 電子メディア化された日本語(特に、インターネット上のテキストデータ)を対象に、言語構造の分析や意味解析などの計量的な言語処理を行う手法を構築する、ことである。そして、(3) 外来語を始めとする語彙の用例を収集・分析する、(4) (3)を日本語コーパス(コンピュータで処理可能な電子化テキスト)として活用する研究を行うことを目的としている。

### 3. 研究の方法

(1) WWWを用いた大規模日本語コーパスの構築とその有効性の検討 本研究では、インターネット上の膨大な日本語テキストを対象に、テキストの自動収集と整形、書誌情報の付加、単語区切りや品詞などの言語的標識の自動付与などを行い、日本語コーパスを構築する手法を考案する。

(2) 文章理解支援への応用 日本語教育の分野では、入門期の学習者を対象に使用頻度

の高い語彙を体系的かつ効果的に教授する工夫がなされており、語彙の使用頻度は日本語教材で使用する語彙の選定および提示順序を決定するための客観的基準となり得る。そこで、本研究では、上記(1)で構築された日本語コーパスを用いて、語彙の使用頻度を集計し、その調査で得られた知見に基づいて教材開発を実施することを通じて、日本語学習者を対象に文章理解を支援する環境を整備する。

(3) 新出「外来語」の分析 日本語コーパスから外来語を抽出し、新出「外来語」の使用頻度を分析する。こうして得られた成果は、新しい外来語へのなじみが薄いといわれている高齢者や日本語学習者に対する日本語理解を促すための教材作成時に必要な基礎的資料となる。また、その資料は、日本語母語話者に対する文章理解を支援する環境を構築するために役立つものにもなり得る。

### 4. 研究成果

本研究では、WWWを日本語コーパスとして活用し、研究の対象とする際の諸問題についても考察した。その結果、(1) サンプルに偏り(バイアス)があること、(2) 執筆者の架空の人格設定が存在すること、(3) 記述内容の妥当性をどのように保証するのか、(4) 同一人物による多数の投稿がみられること、(5) 記述内容の修正や更新などのデータ管理が個人レベルで行われているものが多数存在すること、(6) 不正な文章コピーや他サイトからの引用がみられること、などが考慮すべき点として挙げられた。このように、サンプルのバイアスは残るが、膨大な量の個人発のデータが蓄積できるようになったインターネットは知的資源として研究で利用できるものとなった。今後は、WWWを用いた日本語コーパスのさらなる充実とその応用に向けた研究に積極的に取り組むべきである。

### 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 13 件)

[1] Hironari NOZAKI, ほか 3 名 1 番目 (2010) Study of IT Terms Used in Non-Vocational High School Information Technology Class Textbooks: Toward Corpus-Based Lexical Studies and Sentence Comprehension, 愛知教育大学教育実践総合センター紀要, 13,

59-66, 査読無

[2] 戸田和幸,野崎造成(2010) 「他者意識・自己防衛」を意識した情報モラル教育に関する考察 情報化社会を生きる力との関連, 愛知教育大学教育実践総合センター紀要, 13, 45-49, 査読無

[3] 朴延子,野崎造成,ほか3名2番目(2009) メディア接触と日本語の学習ストラテジーに関する調査 - 中国語および韓国語の母語話者を対象に -, 教育メディア研究, 16(1), 65-75, 査読有

[4] 三浦丈司,野崎造成,ほか3名2番目(2009) 留学生のためのプログラミング言語(HSP)学習支援教材の開発, 教育システム情報学会誌, 26(1), 36-41, 査読有

[5] Tajika Hidetsugu, Hotta Chie, Neumann Ewald, Nozaki Hironari(2009) Episodic memory inhibition and spreading activation: Further evidence of episodic inhibition using the think/no-think task, 神戸親和女子大学研究論叢, 第42号, 55-64, 査読無

[6] Nozaki Hironari, ほか2名1番目(2009) Various Problems Concerning the Construction of a WWW Japanese-Language Corpus The Current State and Future Prospects of Japanese-Language Corpus Research, 愛知教育大学教育実践総合センター紀要, 12, 117-123, 査読無

[7] 戸田和幸,野崎造成(2009) 学校教育におけるネット社会を生きる力の基礎を築く学習の追究 本名・匿名の二つの立場と情報モラル教育の関連, 愛知教育大学教育実践総合センター紀要, 12, 125-130, 査読無

[8] UMEDA Kyoko, TAKITO Shinsuke, NOZAKI Hironari, EJIMA Tetsuro(2008) A proposal for metadata to express students' "presence" in digital photos: Use in the determination of equality in groups of photos, Educational technology research, 31(1・2), 97-106, 査読有

[9] 齋藤ひとみ, 富安伸之, 野崎造成(2008年2月) 高校との地域連携プロジェクトを活用した情報科教育の試み, 愛知教育大学教育実践総合センター紀要, 11, 133-139, 査読無

[10] 野崎造成, ほか4名1番目(2007) 幼稚園・保育園におけるコンピュータ利用教育の現状 - 日本(愛知県)と中国との比較研究 -, 教育メディア研究, 14(1), 41-48, 査読有

[11] Tajika Hidetsugu, Nakatsu Narao, Nozaki Hironari, Neumann Ewald, Maruno Shunichi (2007) Effects of self-explanation as a metacognitive strategy for solving mathematical word problems, Japanese Psychological Research, Vol. 49, No. 3, 222-233, 査読有

[12] 横山詔一・エリク=ロング(2007) 異体

字の単純接触効果に関するロジスティック回帰分析: コーパス4種と最尤推定法を用いた検討, 計量国語学, 26巻1号, 19-30, 査読有  
[13] 江島徹郎, 梅田恭子, 野崎造成, 平田賢一(2007年) 子ども・保護者・教員が相互にBlogを活用して情報を伝え合う力を育成する実践研究, 教育メディア研究, 13(2), 45-53, 査読有

〔学会発表〕(計17件)

[1] 野崎造成, 多鹿秀継(2010年, 印刷中) 他者意識と受信絵文字量が送信絵文字量に及ぼす影響, 日本心理学会第72回大会, 大阪大学

[2] 田中佳子, 野崎造成, ほか2名2番目(2009年11月7日) 日本語力を測る・育てる - 測りやすい力と育てる力 -, ITK-FUKUOKA2009 Interactive Technology in Education (ITK: Interaktiivinen Tekniikka Koulutuksessa) 日本リメディアル教育学会九州・沖縄支部合同講演会論文集, 23-26, 福岡大学

[3] 野崎造成, 田中佳子(2009年9月1日) 日本語検定試験の分析~高校生の日本語運用能力の経年変化~, 日本リメディアル教育学会第5回全国大会発表予稿集, 9-10, 千歳科学技術大学(北海道千歳市)

[4] 田中佳子(2009年9月1日) 日本語検定試験の分析~語彙知識をどう測るか~, 日本リメディアル教育学会第5回全国大会発表予稿集, 11-12, 千歳科学技術大学(北海道千歳市)

[5] 野崎造成, 多鹿秀継(2009年8月26日) インターネット依存とLocus of Controlの関係について, 日本心理学会第73回大会, 1E061, 立命館大学(京都府)

[6] 戸田和幸, 野崎造成(2009年8月21日) 学校教育におけるネット社会を生きる力の基礎を築く学習の追究 - 本名・匿名の二つの立場と情報モラル教育の関連 -, 教育システム情報学会第34回全国大会講演論文集, 490-491, 名古屋大学, [研究奨励賞受賞論文]

[7] 近藤衣里子, 野崎造成, ほか2名2番目, (2009年7月4日) 読み書きが苦手な児童の漢字学習支援--デジタルペンを使用した漢字誤り分析--, 教育システム情報学会研究報告, 24(2), 23-27, 北海道大学

[8] 大原由紀子, 馬場由佳里, 戸田和幸, 中津植男, 野崎造成, 江島徹郎, 梅田恭子(2009年3月6日) 教科教室型における理想的な数学科教室の提案, 2008年度JSiSE学生研究発表会, 95-97, 名古屋大学

[9] 野崎造成, 横山詔一, ほか3名1番目, (2009年7月16日) 新聞における「外来語」使用の経年変化の分析, JSAA-ICJLE 2009年度研究大会, Session7, Mathews308, ニューサウ

スウェールズ大学(豪州,シドニー市)

[10] 横山詔一,ほか3名1番目,(2009年1月23日)平成明朝体のデザイン変異に関する共時的認知分析,情報処理学会研究報告2009-CH-81,33-40,キャンパスプラザ京都

[11] 野崎造成,多鹿秀継(2008年9月20日)大学生におけるテレビCMと新聞広告の効果・機能,日本心理学会第72回大会,2AM091,北海道大学

[12] 田中雅也,野崎造成,梅田恭子,江島徹郎(2008年7月26日)ネット討論による授業実践の報告とその考察,教育システム情報学会研究報告,23(2),3-8,愛知教育大学

[13] 野崎造成,三浦丈司,田中佳子,江島徹郎,梅田恭子(2008年7月11日)留学生のためのプログラミング言語(HSP)学習支援教材の開発,日本語教育学世界大会2008予稿集3,302-305,釜山外国語大学(韓国)

[14] 福上慎吾,野崎造成,江島徹郎,梅田恭子(2007年10月20日)ディベートによる情報活用の実践力の育成 - 高校生を対象とした授業実践 -,教育システム情報学会研究報告,22(3),22-26,名古屋大学

[15] 野崎造成,多鹿秀継(2007年9月19日)情報教育で活用されるメディアに対する高校生・大学生の認知,日本心理学会第71回大会,2AM056,589,東洋大学

[16] 朴延子,野崎造成,江島徹郎,梅田恭子,平田賢一(2007年7月14日)メディア接触が日本語の学習ストラテジーに及ぼす影響--中国語および韓国語の母語話者を対象に--,教育システム情報学会研究報告,22(2),3-7,千歳科学技術大学

[17] 戸田和幸(2007年4月28日)学校教育におけるネット社会を生きる力の基礎を築く学習について,教育システム情報学会東海支部2007年度第1回研究会,名古屋市

〔図書〕(計4件)

[1]Shoichi Yokoyama and Haruko Sanada(2009) Logistic regression model for predicting language change, Studies in Quantitative Linguistics 5, Issues in Quantitative Linguistics, 176-192, RAM-Verlag, Germany

[2]横山詔一(2009)「文字はココロの眼と指で読む」『日本語学』12月号,40-48,明治書院(東京都)

[3]多鹿秀継 編著、あいり出版、学習心理学の最先端 --学びのしくみを科学する--, 2008、186 (野崎造成 第3章 26-37頁執筆)  
(中津 植男 第14章 160-171頁執筆)

[4]横山詔一(2008)「言語の単純接触効果」『単純接触効果研究の最前線』,114-127,北大路書房(京都市)

〔産業財産権〕

出願状況(計0件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
出願年月日:  
国内外の別:

取得状況(計0件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
取得年月日:  
国内外の別:

〔その他〕  
ホームページ等

## 6. 研究組織

### (1) 研究代表者

野崎 造成 (NOZAKI HIRONARI)  
愛知教育大学・教育学部・准教授  
研究者番号: 80275148

### (2) 研究分担者

中津 植男 (NAKATSU NARAO)  
愛知教育大学・教育学部・教授  
研究者番号: 90133131

江島 徹郎 (EJIMA TETSURO)  
愛知教育大学・教育学部・准教授  
研究者番号: 10335078

梅田 恭子 (UMEDA KYOKO)  
愛知教育大学・教育学部・准教授  
研究者番号: 70345940

齋藤 ひとみ (SAITO HITOMI)  
愛知教育大学・教育学部・講師  
研究者番号: 00378233  
(H20 H21: 研究分担者)

平田 賢一 (HIRATA KENICHI)  
岡崎女子短期大学・幼児教育学科・教授  
研究者番号: 20024060  
(H20 H21: 連携研究者)

多鹿 秀継 (TAJIKI HIDETSUGU)

神戸親和女子大学・発達教育学部・教授  
研究者番号：20024060

横山 詔一 (YOKOYAMA SHOICHI)  
国立国語研究所・理論・構造研究系・教授  
研究者番号：60182713

(3) 連携研究者

( )

研究者番号：

(4) 研究協力者

田中佳子 (TANAKA YOSHIKO)  
日本工業大学・工学部・講師

戸田和幸 (TODA KAZUYUKI)  
名古屋市立砂田橋小学校・教諭