

領域略称名：日本語コーパス
領域番号：125

平成23年度科学研究費補助金
「特定領域研究」に係る研究成果等の報告書

「代表性を有する大規模日本語書き言葉コーパスの構築：
21世紀の日本語研究の基盤整備」

(領域設定期間)
平成18年度～平成22年度

平成23年6月

領域代表者 大学共同利用機関法人人間文化研究機構国立国語研究所・
言語資源研究系・教授・前川喜久雄

目 次

1. 研究領域の目的及び概要	1
2. 研究領域の設定目的の達成度	2
3. 研究領域の研究推進時の問題点と当時の対応状況	3
3. 1 著作権処理	3
3. 2 国立国語研究所の移管問題	3
4. 主な研究成果	4
4. 1 研究項目A01「コーパスの構築」	4
4. 2 研究項目B01「コーパスの評価」	5
4. 3 公募班	7
5. 研究成果の取りまとめの状況	10
6. 研究成果の公表の状況	11
6. 1 主な論文等一覧	11
A) 査読済論文	11
B) 無査読論文	13
C) 著書	15
D) 口頭発表	15
E) 特定領域で刊行した報告書	18
6. 2 ホームページ	20
6. 3 公開発表	21
6. 4 「国民との科学・技術対話」	23
7. 研究組織と各研究項目の連携状況	24
7. 1 研究組織	24
7. 2 項目間の連携状況	26
8. 研究費の使用状況	28
9. 当該学問分野及び関連学問分野への貢献度	28
9. 1 シンポジウム・特集・報道	28
9. 2 『現代日本語書き言葉均衡コーパス』モニター版利用者	30
10. 研究計画に参画した若手研究者の成長の状況	30
11. 総括班評価者による評価の状況	30

1. 研究領域の目的及び概要

研究領域名：代表性を有する大規模日本語書き言葉コーパスの構築：21世紀の日本語研究の基盤整備

研究期間：平成18年度～平成22年度

領域代表者所属・職・氏名：大学共同利用機関法人人間文化研究機構国立国語研究所・言語資源研究系・教授（系長）・前川喜久雄

補助金交付額：

（単位：千円）

平成18年度	平成19年度	平成20年度	平成21年度	平成22年度	総額
136,800	203,900	171,300	139,600	131,800	783,400

※中間評価でのA評価を反映して平成21年度と22年度に研究費の追加配分があったが、上の数字には含めていない。

近年、世界の言語研究の著しい特徴として、電子的に蓄積された大規模な言語資料、すなわち言語コーパス(corpus)を用いた研究の隆盛が指摘できる。数億語規模のデータをコンピュータ上で解析することにより、言語の実態に正確に基づいた研究が可能になり、その結果、従来のような研究者の内省に頼った研究では把握が不可能であったり困難であったりした現象が分析の俎上にのぼるようになってきた。

コーパスはまた様々な応用研究にも有益である。数例を挙げれば、母語話者用辞書および学習者用辞書の双方における用例と意味記述の質的向上、国語の義務教育において教育すべき語の範囲（教育用語彙）の策定と学年配当、さらに常用漢字表の見直し作業の基礎資料としての利用などである。

このようにコーパスは言語研究の枠を超えた幅広い学術分野で科学的営為を支える知的資源となる可能性がきわめて高いのだが、残念なことに我が国における言語コーパスの整備状況は、欧米のみならず、韓国、台湾などのアジア諸国に比べても出遅れている。わけても日本語の多様な全体像を偏りなく反映した均衡コーパス（すなわち代表性—representativeness—を備えたコーパス）が構築されていない点で、先行する諸外国に著しく水をあけられている。

本研究領域にはふたつの大きな目標がある。ひとつは、現代日本語のコーパス言語学的研究の基盤を整備するために、大規模な現代日本語書き言葉の均衡コーパスを構築することである。本領域におけるコーパス構築は、国立国語研究所の近現代日本語コーパス整備事業であるKOTONOHA計画と連携して実施する。両者が協力して1億語を超える規模の『現代日本語書き言葉均衡コーパス』を構築するが、分担関係として、本領域では書籍に用いられた現代語の書き言葉を対象とする5000万語規模の書き言葉コーパスを構築し、KOTONOHA計画では、それ以外（新聞、雑誌、ウェブ文書、その他）の5000万語を構築する。サンプルには著作権処理を実施して、誰でも自由に利用できるコーパスとして公開する。

本領域のもうひとつの目標は、構築途上のコーパスを様々な領域で利用することによってコーパス日本語学の可能性を探ることである。狭義の言語学だけでなく、国語教育、日本語教育、辞書編集、自然言語処理などの幅広い領域で活用の可能性を探る。このように本特定領域研究の成果は、今後長期間にわたって広い範囲の日本語研究に影響をおよぼすと予想される。本領域の正式名称の副題部分にはそのような意気込みがこめられている。

本領域にはふたつの研究項目（A01とB01）を設定し、それぞれ上記の第1、第2の目標達成をめざした研究を実施した。A01には3班、B01には5班の計画研究班を設置し、これに加えて第2年度からは公募研究4班が活動した。

2. 研究領域の設定目的の達成度

第1の目標である均衡コーパスの構築はほぼ全面的に達成したと考えている。『現代日本語書き言葉均衡コーパス』には現在約1億500万語（短単位）相当のテキストが蓄積されているが、そのうち6000万語以上が本領域で作成した書籍のデータであり、目標とした5000万語を上回るデータを作成した。『現代日本語書き言葉均衡コーパス』は2011年3月に全文検索可能な形でオンライン公開しているが、8月にはコーパス全体に短単位と長単位による形態論情報を二重に付与したデータをウェブ上で公開する予定であり、現在その最終準備を進めている。

形態論情報の解析精度についても、短単位の場合、コーパス全体では目標とした98%を上回る精度を達成している。長単位の解析は現在も結果を調整中であるため、正確な精度を報告することはできないが、書籍ないし雑誌のテキストでは95~97%の精度に達している。以上の成果は主にデータ班と電子化辞書班の緊密な連携によって実現されたものである。

『現代日本語書き言葉均衡コーパス』の形態素解析のために電子化辞書班が構築を進めた形態論情報解析用電子化辞書UniDicは一般に無償公開しており3755名以上の登録ユーザーがいる。

第2の目標であるコーパス日本語学の可能性の探索は、その性質上、第1の目標のように明瞭かつ定量的な到達目標に沿って達成度を評価することはできない。ここではアウトプットの量によって評価を試みる（詳細は6節参照）。本領域の5年間の活動の成果として、査読論文192篇（そのうち71篇は英語で執筆されfull paperで事前査読された国際学会予稿集論文）、無査読論文224篇、著書10篇、口頭発表541篇、報告書52篇（12,178頁）が公開されている。本領域に参加した研究者数は常時約80名であるから、平均すると各員が査読論文2.4篇、無査読論文2.8篇、口頭発表6.8件を発表したことになる。正確なデータが存在するわけではないが、関連学界の成果公表の平均的なペースを大きく上回っていることは間違いない。

本領域の活動が関連諸学会に広く大きな影響を及ぼしたことは、日本語学会、日本言語学会、英語コーパス学会、人工知能学会が本領域の活動に注目したシンポジウムや学会誌特集号を企画したこと、論文誌「日本語科学」と商業誌「国文学解釈と鑑賞」が本領域の活動を集めたことなどによって立証できる。領域代表者も国内外で基調講演、招待講演、シンポジウムパネル発表、インタビュー等を20回行った（詳細は9節参照）。

本領域では、2007年度以降、『現代日本語書き言葉均衡コーパス』のために収集されたテキストのうち、著作権処理が終了したものをウェブ上で全文検索可能な形でデモ用に公開してきた。このデモサイトには通算で14万件を超えるアクセスがあり、これを主要なデータ源として執筆された論文も公開されている。

やはり2007年度には構築途上の『現代日本語書き言葉均衡コーパス』のテキストデータ約4000万語を『モニター版』として領域外の研究者に研究目的で無償公開し、2009年にも約7000万語を第二次モニター公開した。これらのモニター版には通算で900件以上の利用申請があった。

モニター版による研究成果発表の場を設けるために、各年度末に開催した公開研究成果発表会に領域外の研究者も応募できる公開サテライトセッションを設置した。このセッションには3年間で約70件の応募があり、コーパス日本語学の普及に貢献した。

さらに上述のUniDicは、自然言語処理の領域も含めて日本語解析用辞書の新しい標準として広範に利用されている（9.2節参照）。

このように、本領域の活動がコーパスによる日本語研究を活性化したことは明らかであり、本領域の第2の目標も十分に達成されたものと考えられる。

3. 研究領域の研究推進時の問題点と当時の対応状況

3. 1 著作権処理

本領域の推進時に生じた大きな問題はふたつあった。ひとつはコーパス公開のための著作権処理である。この問題の所在は予め想定しており、領域申請書の段階から総括班が著作権処理の窓口となってデータ班と共同で作業にあたることを決めていた。『現代日本語書き言葉均衡コーパス』の著作権処理作業量は膨大であり、5年間で26000件以上のサンプルの著作権処理を実施した。この作業の過程で痛感されたのは、現在の著作権法ではこのような大量の著作権処理が想定されておらず、法律がネット社会に対応できていないことであった。この問題はひとり言語資源学に限らず、情報処理研究の広い範囲に関係する問題であるため、著作権法の改正の必要性が痛感された。

幸い関係諸方面からはたらきかけもあって、文化審議会著作権分科会法制問題小委員会においてネット上のテキストを研究用に利用する場合の著作権に関する権利制限の見直しが審議されることとなり、本領域代表者も平成20年7月に参考人として意見を述べる機会を得た（当日の詳細な議事録がhttp://www.bunka.go.jp/chosakuken/singikai/housei/h20_05/gijiroku.htmlに公開されている）。その後平成22年1月1日に著作権法が改正されたことは周知の通りである。

この改正は一部で「平成の大改正」と呼ばれているが、当初導入が予定されていた「日本版フェアユース規定」の導入が見送られたため、本領域のかかえる問題を全面的に解決するものとはならなかった。しかし現代の言語研究と著作権法が密接に関係していることを著作権法の専門家に訴えることができたのはひとつの進歩であった。

『現代日本語書き言葉均衡コーパス』の著作権処理は、上記の改正後も従来と同じ方式で継続し、本稿執筆の時点で9000万語相当のテキストの公開許諾が得られている。著作権処理は本領域終了後も国立国語研究所の業務として継続しており、最終的に1億語全体の許諾を得るための努力を続けている。

3. 2 国立国語研究所の移管問題

第二の問題は本領域項目A01の活動の中心となっていた国立国語研究所の移管問題である。行政減量・効率化有識者会議の答申をうけて自民党福田内閣は2007年12月24日の閣議で独立行政法人である国立国語研究所を廃止し、大学共同利用機関法人に移管することを決定した。この答申では国立機関によるコーパス構築の必要性が全面否定されており、民間で実施すべきものとされていた（詳しくはhttp://www2.ninjal.ac.jp/kikuo/ikan_record.html参照）。そのため本領域と連携して『現代日本語書き言葉均衡コーパス』の書籍以外の部分を構築していた国立国語研究所のKOTONOHA計画の存続が危ぶまれる事態が生じた。

その後政府から新研究所の具体的なありかたの検討を委ねられた科学技術・学術審議会は2008年7月の報告で新研究所を人間文化研究機構に設置すべきこととし、さらに新研究所が「当面、特に重点を置いて推進する必要がある研究分野」のひとつとして言語資源研究を認めたので、上述の危機を脱することができた。しかし2008年から2009年にかけては、若手研究員の処遇を巡って人間文化研究機構との交渉があり、また新研究所の組織設計にともなう膨大な事務作業も発生し、領域代表者を含む本領域の中心的なメンバー数名が、それらの対応に忙殺されることとなった。ほぼ1年間におよぶ移管準備期間に『現代日本語書き言葉均衡コーパス』の構築作業をあらかじめ予定どおりに推進できたのは、データ班および電子化辞書班メンバーの強い使命感によるものであった。

国立国語研究所は2009年10月1日に人間文化研究機構傘下の1研究所として再出発した。新研究所には言語資源研究系に加えて言語資源開発業務のためにコーパス開発センターが設置されることとなり、本特定領域研究の成果を継承して一層の発展をめざすための組織が整備される結果となった。

最後に、領域運営上の細々とした問題は各年度とも4回開催した総括班会議で検討した。またコーパス構築上の問題については、ほぼ月例で進捗状況確認のためのミーティングを開催（通算59回）して問題点の把握と解決に努めた。

4. 主な研究成果

4. 1 研究項目 A01「コーパスの構築」

A1)データ班

- (1) 『現代日本語書き言葉均衡コーパス』の6割以上を占める書籍コーパス（約6,500万語）の構築を完了した。『現代日本語書き言葉均衡コーパス』は2011年8月の一般公開の準備を進めている。
- (2) 構築と平行して形態素解析用辞書UniDicに搭載する短単位データベースを電子化辞書班と連携して整備した。辞書の見出しに当たる語彙素数は約21万である。
- (3) 形態素解析結果を手で修正し精度を99%に高めたコアデータを完成させた。コアデータには、短単位、長単位、文節の3種類の言語単位の境界情報を付与した。語数は短単位で約110万である。
- (4) 著作権処理はすべての書籍サンプルについて連絡先を調査し、連絡先が判明したものについては許諾依頼を行った。
- (5) コーパスの構築に関するノウハウをまとめた報告書を17冊（総計2808頁）刊行した。
- (6) 形態素情報検索用ウェブアプリケーション「中納言」のプロトタイプを開発した（図1参照）。

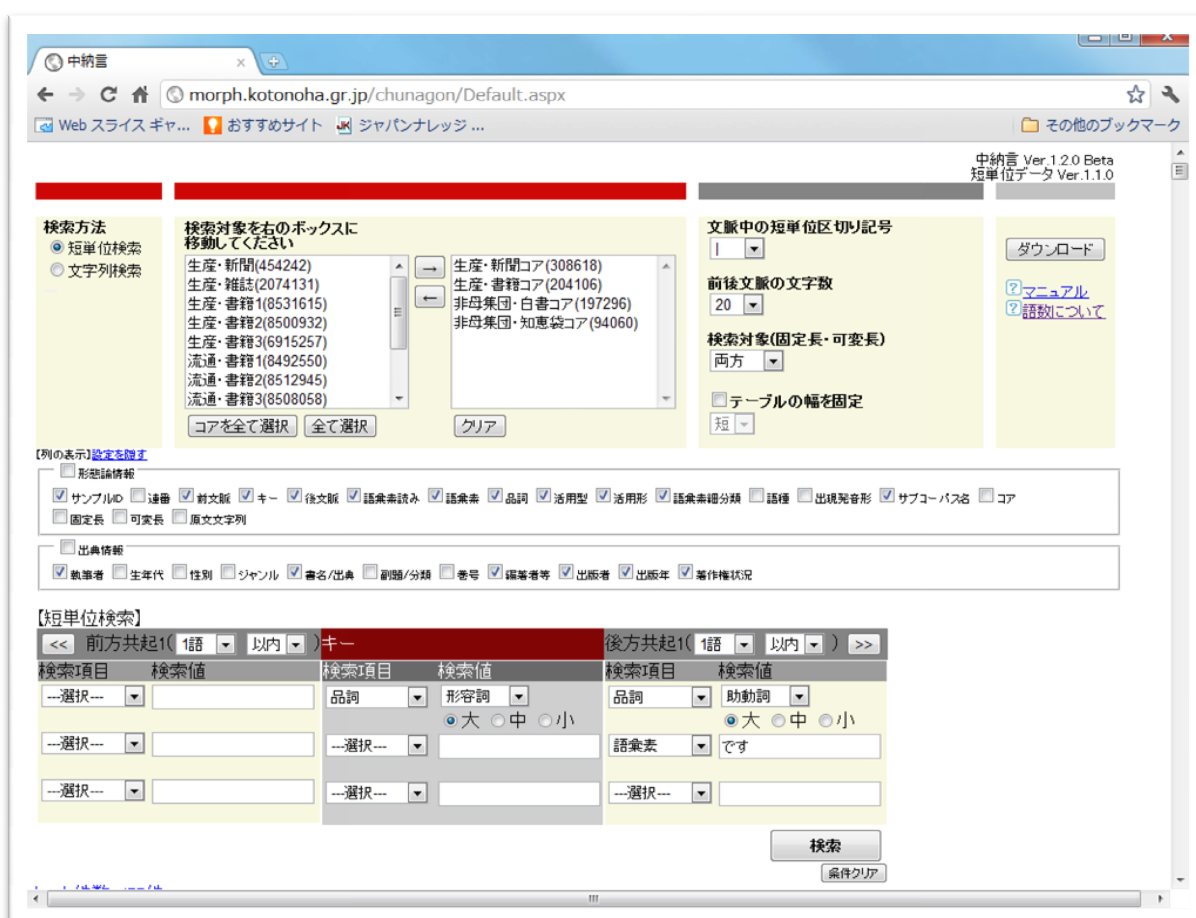


図1: 「中納言」の検索条件指定画面（プロトタイプ）

A2)電子化辞書班

- (1) 以下の特徴を持つ形態素解析用電子化辞書を設計・開発した。
「短単位」という揺れがない斉一な単位で設計
語彙素・語形・書字形・発音形の階層構造を持ち、表記の揺れや語形の変異にかかわらず同一の見出しを与えることが可能
アクセントや音変化の情報を付与でき、テキスト音声合成などに利用可能
- (2) 辞書データベースを国立国語研究所内で構築しながら、形態素解析システムChaSen・MeCab用

辞書を随時公開し、最終的に語彙素約21万語・書字形約33万語の規模と、品詞認定約98.9%・語彙素認定約98.6% (MeCab版) の解析精度を達成できた。最終年度には、辞書データベースをXMLファイル群として記述し、ユーザーがカスタマイズ可能な辞書作成環境を提供する新しい方式でUniDic2を設計・開発した。

- (3) 中・長単位解析システムを含む、形態素解析の後処理ツール群を作成し、多様な目的に供することができた。

A3) ツール班

- (1) 『現代日本語書き言葉均衡コーパス』に代表される日本語書き言葉コーパスに対する様々な言語情報のアノテーションを自動的に行う言語解析ツールの開発およびアノテーションの誤り修正やアノテーションを施されたコーパスの柔軟な利用や管理を行うためのコーパスツールの開発を行った。具体的には、形態素解析、文節分かち書き、文末判定、係り受け解析、並列構造解析、固有表現認識、述語項構造解析、照応・共参照解析、事象間時間関係解析などの言語解析ツールを開発した。
- (2) アノテーションを手動で行う作業を支援するため、また、自動的にアノテーションされたコーパスに対する誤り修正などの追加のアノテーション作業を行うためのコーパス構築支援ツールを開発した。具体的には、形態素解析、文節分かち書き、係り受け解析、並列構造解析が施されたコーパスを対象に、コーパスの検索、検索結果の表示、各種統計処理、アノテーションの修正作業などを支援する「茶器」、アノテーション作業を文書のある領域 (セグメント) の範囲指定およびラベル付けとセグメント間の関係のラベル付け、という一般的な作業として定義した汎用アノテーションツール「Slate」、そして複数のコーパスツールの相互運用を行うためのツール (名称未定) を開発した。

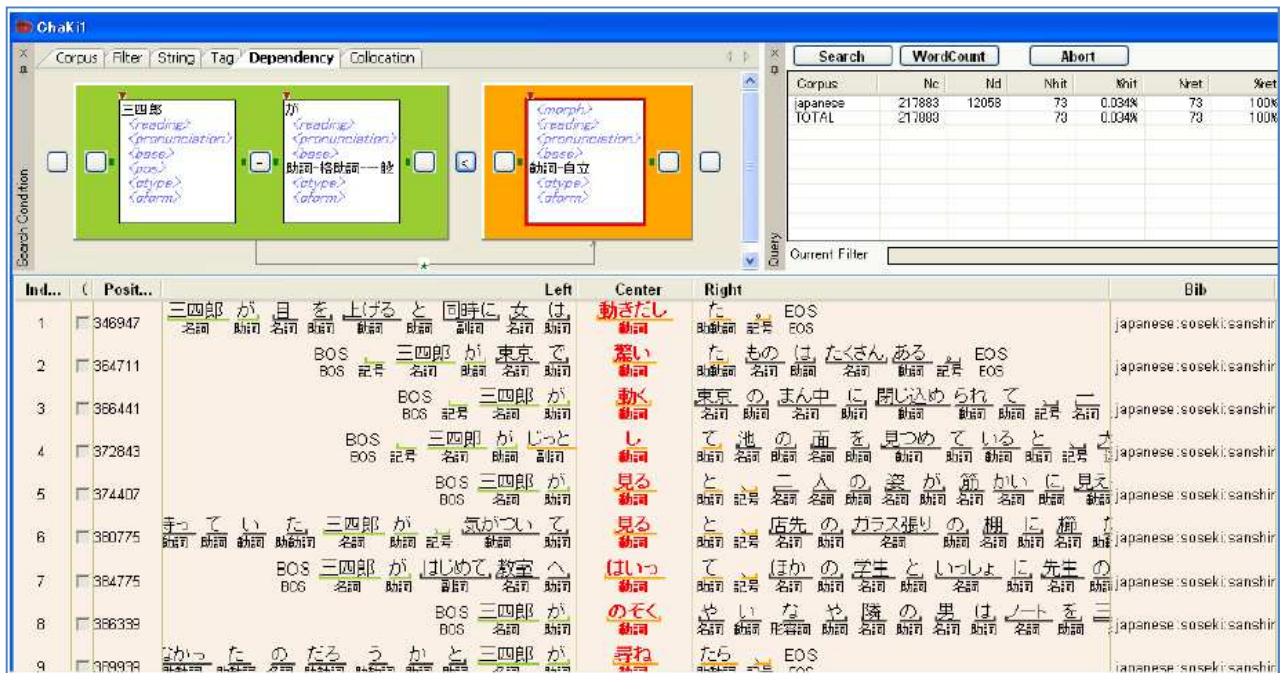


図2: 「茶器」の係り受け関係検索のKWIC表示画面

4. 2 研究項目 B01 「コーパスの評価」

B1) 日本語学班

日本語研究におけるコーパス利用の価値と可能性を明らかにし、コーパスに基づく日本語研究の振興を図ることを主たる目的として活動を行った。5年間の研究内容は多岐にわたるが、研究課題名にもある次の2つの観点に重点を置いて研究を進めた。

- (1) ひとつは「日本語研究の精密化」であり、コーパスから得られる豊富な用例を利用することによ

り、従来少数の用例や内省に依存して行われてきた種類の日本語研究を精密化し、より実証的なものにする可能性を追求した。具体的には、複合辞やコピュラの考察、語句の意味分析などに取り組んだ。

- (2)もうひとつは「新しい研究領域・手法の開発」であり、コーパスの特性を生かすことにより、旧来の言語研究資料では実現できなかった様々な種類の研究への道を拓く可能性を追求した。具体的には、現代日本語の通時変化の解明、コロケーションの分析、画像や音声を含むマルチメディア・コーパスの可能性などに取り組んだ。

B2)日本語教育班

BCCWJを日本語教育に活用する方法の開発を目標として、①日本語教材コーパスの作成と分析、②コーパスを活用した日本語教材作成法の開発、③日本語教育のためのコーパス活用ツールの開発という3つの課題を立てて活動した。

- (1)上記①については、初級・中級向け日本語総合教科書と読解教材48種の入力とタグ付けを行い、現在著作権処理を継続中である。
- (2)上記②については、BCCWJとGoogle刊行のN-gramファイルを利用したコロケーションリスト作成、BCCWJ領域内公開データ2009年度版を用いた機能語データベース構築、同データを用いた語彙頻度分布調査に基づく日本語学習語彙表作成などがある。
- (3)上記③については、BCCWJを用いた語彙統計情報データベース作成、語彙パターンを柔軟に検索するための検索ツール試作などがある。

下表に成果報告書DVDに収録した日本語教育班関係のコンテンツを示す。

分担者	コンテンツ
全員	平成18年度～22年度の研究成果報告書
小林・小西	BCCWJを用いた初級シラバスの再評価
滝沢	言語研究のためのテキスト処理入門
滝沢	BCCWJ2009Ngrams
砂川・清水・奥川・千葉	BCCWJによる機能語データベース（スタンドアロン版）（Ver. 0.9.1b）
投野・本田	BCCWJサブコーパス10種の評価用語彙頻度・分布統計
千葉	BCCWJによる語彙情報データベース（Ver. 0.6）

B3)言語政策班

- (1)『現代日本語書き言葉均衡コーパス』をもとに、ジャンルやレベルで分類した語彙表と漢字表を作成した。また、小・中・高校で用いられている全教科を対象とした教科書コーパスをもとに語彙表と漢字表を作成し、社会の語彙・漢字と、学校のそれとを対照できるようにした。これらを基礎資料として、国語政策と国語教育の諸課題にコーパスを導入する研究を進めた。
- (2)国語政策については、専門用語を専門度や重要度などで序列化し、専門情報を非専門家に分かりやすく伝える工夫の提案に役立てる研究を進めたほか、漢字のジャンル別・語別の頻度表を、「常用漢字表」の見直しなど漢字政策の検討材料に用いる研究を行った。
- (3)国語教育の分野では、教科やレベルの観点から重要語彙や重要漢字を分類し、教科を横断する指導や、生徒の発達段階に応じた指導のあり方を研究したほか、作文を支援するツールとして、コーパスに基づく語彙文例集を活用する方法を検討した。国語教師対象のワークショップも実施して、コーパスをこの分野に普及させるための課題も探った。

B4)辞書編集班

- (1)名詞と動詞のコロケーション研究では、BCCWJのサイズが研究資料として不十分であるとい

う結論に達した。そのためWWW を活用する手段について研究をすすめた。コロケーションを数値的に判断する方式はある程度確立したものの、実際には大量の用例の中に意図しない用例が相当量混じっており、半自動的なコロケーション辞書作成は無理であることもわかった。

- (2) 複合辞の研究では、BCCWJ を基にした複合辞資料集ができあがった。これは、n-gram の手法で自動抽出した資料から頻度・スコアで複合辞候補を絞り込み、さらにそれを人手でチェックしてまとめたものである。その資料を用いて、各種の文法研究などを展開することができた。
- (3) 国語辞典の意味記述、意味区分、文法事項の認定などに対し、アンケートおよびBCCWJ を利用した実態調査を行い、記述の妥当性を検証した。この成果にもとづいて、一部の国語辞典では記述の修正が行われた。
- (4) 動詞の格情報については、コーパスにおける使用実態を実際に岩波国語辞典第七版に反映させた（主に例文として記載）。動詞の格情報についてもオノマトペについても、コーパスの情報を踏まえた辞書を試作した。基礎的な動詞の格情報の記述にはBCCWJ がかなり有効だが、オノマトペの記述には不十分であることが明らかになった。文法的事項と語彙的事項の差と言えるかもしれない。

B5)言語処理班

- (1) 語義タグ付コーパスの構築では『現代日本語書き言葉均衡コーパス』コアデータに対して、岩波国語辞典中の語義の区分に基づき人手で語義を付与する作業を行った。
- (2) 新しい語義曖昧性解消タスクでは、語義曖昧性解消に関する評価型ワークショップである Semeval-2 (<http://semeval2.fbk.eu/Semeval2.html>) に『現代日本語書き言葉均衡コーパス』を用いた語義曖昧性解消の評価型タスクを提案し採択された。現時点では領域内5 グループ、領域外の国内4 グループ、領域外の海外4 グループが上記(1)で構築した語義タグ付コーパスを利用した研究を実施している。
- (3) 代表性のあるコーパスを用いた語義曖昧性解消では、ソースデータとターゲットデータの組み合わせごとに効果的な領域適応手法を自動的に選択する手法の開発を行った。領域適応手法の自動選択はソースデータとターゲットデータの性質に関する情報を元に決定木学習を用いて行う。自動的に選択された領域適応手法を用いることで語義曖昧性解消の性能が有意に向上することが確認された。
- (4) 半教師ありクラスタリング手法の開発と多義性解消への適用では、クラスタリング時に教師情報を部分的に利用する半教師ありクラスタリング手法を開発した。
- (5) コーパスからの新語義の発見では、まず対象単語を含む用例をコーパスから収集し、同じ語義を持つ用例をまとめたクラスタを作成する。用例は4 種類の特徴ベクトルで表現する。クラスタリングの際にはこれらの特徴ベクトルのいずれかの類似度が高い用例をまとめてクラスタを作成する。1 種類の特徴ベクトルを用いる先行研究と比べると、複数の特徴ベクトルを同時に用いることで語の類似性を様々な観点から評価できる点に提案手法の特徴がある。
- (6) 特異用例の検出にかかわる提案手法は2つの検出手法からなる。第1 の手法はLOF を教師付きの枠組みに拡張したものである。第2の手法は教師データから語義識別の分類器を学習し、各データの語義を推定する。推定された語義のクラスタとデータとの距離関係から外れ値かどうかを判定する。提案手法では第1 の手法により外れ値の候補を取り出し、第2 の手法でその候補を選別する。
- (7) 同時クラスタリングを利用した動詞類義語獲得については、動詞の類語をテキスト中の係り受け関係から獲得する手法として、同時クラスタリングがベクトルベースの動詞だけのクラスタリングに対して有効であることを明らかにした。

4. 3 公募班

本領域では平成18年と20年に公募班を募集した。いずれも4班が採択されたが、うち2班（下記

のC3とC4)は同一の研究班が2回続けて採択されている。またC2の研究班は平成21年度以降は言語処理班(上記B5)の一部として活動した。

C1)リーダビリティ班(公募班:平成19,20年度)

文章の読みやすさに関する研究をおこなった。その成果として以下の特許を出願した。

特許出願:「文章の読み易さ評価システム及び文章の読み易さ評価方法」(特願:2007-169544平成19年6月27日出願,平成23年5月24日審査請求開始)

日本語の文章の読みやすさを従属変数とし,1文の平均文字数,1文の平均述語数,文章全体の平仮名の割合,文章全体の漢語の割合を独立変数とする重回帰式によるリーダビリティの予測式。平成20年7月1日より特許庁HPに公開されている。

C2)クラスタリング班(公募班:平成19,20年度)

半教師ありクラスタリング手法の開発と多義性解消への適用では,クラスタリング時に教師情報を部分的に利用する半教師ありクラスタリング手法を開発した。

C3)作文支援システム班(公募班:平成19~22年度)

『現代日本語書き言葉均衡コーパス』の日本語教育への応用および日本語学習者からのフィードバックによる評価を目的とし,外国人学習者に効果的と考えられる共起表現習得に焦点をあわせ,コーパス利用によるData Driven Language Learningの考えに基づいて作文支援システム構築を行った。

- (1)『現代日本語書き言葉均衡コーパス』から学習者レベル・学習目的に適切な例文表現を抽出して学習者に理解しやすい提示方法の開発をし,コーパスのバランス性を評価した。後半では学習目的に適切なジャンル別の表示をし,学習者に理解しやすい提示システムを考案し,利用する観点からBCCWJのさまざまなジャンルとともに,新たに収集した学術論文を追加して,学習者の作文支援が有効に機能することを明らかにした。
- (2)さらに学習者作文を収集し,誤用タグ付けをすることで学習者コーパスを構築し,正用コーパスとしてのBCCWJと組み合わせて利用することで特にレジスターという概念を取り入れることで,コーパス利用による学習者支援が有効であることを学習者実験により明らかにし,BCCWJが日本語教育支援に利用できる可能性が大であると評価した。
- (3)また遠隔地の日本語学習者に対する使用実験においても有効であることを示した。

C4)日本語フレームネット班(公募班:平成19~22年度)

- (1)語彙資源「日本語フレームネット」のプロトタイプをフレーム意味論の枠組みでBCCWJデータに意味アノテーションを行うことで構築し,公開した。アノテーション・データのうち語彙アノテーション・データはFrameSQLという検索・表示ツールを介して公開した(<http://sato.fm.senshu-u.ac.jp/jfn23/notes/index2.html>)。
- (2)全文テキストアノテーション・データは,BCCWJコアデータ書籍ジャンル全サンプルに対して行った。『現代日本語書き言葉均衡コーパス』が一般公開され次第全文テキストアノテーションWeb Report上で公開予定である。
- (3)ツールとして,FrameSQL,全文テキストアノテーションWeb Reportの他に,JFN-KWIC検索システムというBCCWJデータ検索システムを開発した。

C5)日本語機能表現班(公募班:平成21,22年度)

日本語の機能表現を構成する表記について,複合辞としての機能的用法と表記を構成する内容語本来の語義を持つ内容的用法との間の曖昧性を解消する手法を確立した。提案手法においては,階層的機能表現辞書「つつじ」(17,000表現を収録)を利用し,階層において下位に位置する派生的表

現について用法が類似するより上位の代表的表現の用例を参照して用法判定を行う。提案方式に基づき、代表的表現の表記の用法判定済み用例集合(約 38,000 例)を参照して派生的表現の表記の用法判定を行うことにより 80%以上の用例の用法を正しく判定できることを示した。また、日本語特許文中の機能表現の英訳において、階層的機能表現辞書の 199 個の意味的等価クラスを用いることにより意味的等価クラスごとに少数の英訳規則を作成する集約的英訳方式を提案しその評価を行った。

C6) 意見情報班（公募班：平成 21,22 年度）

意見情報班では、『現代日本語書き言葉均衡コーパス』（BCCWJ）から Yahoo!知恵袋と書籍、それ以外に国立情報学研究所が提供する NTCIR 意見分析コーパスから新聞記事、また国際会議 ICWSM2009 で提供されたコーパスからブログを主な対象とすることによってさまざまな文書ジャンルやドメインを対象として意見情報の分析を進め、傾向の違いを整理し、情報アクセスに着目した応用を進める上で必要となる意見情報を明らかにした。

また、これらの意見情報の特徴の違いを判別することを目的としてアプレイザル理論に基づく英語の辞書から日本語のアプレイザル辞書を構築し、意見抽出における効果を検証した。検証に当たっては国際評価型ワークショップ NTCIR-8 多言語意見分析タスクを開催し、そこで作成した評価用データセットを利用した。さらに口語表現・会話表現に焦点を当てて、Yahoo!ブログと国会会議録を対象とした意見分析コーパスを作成した。2年間の活動期間を通して作成した意見分析コーパスのサイズを下表に示す。

文書ジャンル	ソース	文書数	総文数	内容	作成年度
Yahoo!知恵袋	BCCWJ	251	1,924	コアデータのうち、主要7カテゴリ	平成21年度
書籍		10	407	評論、随筆など	平成21年度
Yahoo!ブログ		471	6,944	コアデータすべて	平成22年度
国会会議録		14	5,812	本会議録	平成22年度
新聞	NTCIR	780	21,391	NTCIR-6, 7 MOAT	平成21年度
Yahoo!知恵袋	API	16	118	アプリケーション用途	平成21年度
ブログ	ICWSM	80	2294	アプリケーション用途	平成21年度

5. 研究成果の取りまとめの状況

- (1) 本領域の活動期間中に研究成果発表会（毎年度9月と3月の2回開催）の予稿集を合計12冊、各班の報告書を32冊刊行した。
- (2) 平成23年8月2,3日に東京で『現代日本語書き言葉均衡コーパス』完成記念講演会を開催する。
- (3) 本領域で得られた日本語コーパスの構築と分析に関する知見を「講座日本語コーパス」（全8巻,合計2000頁,朝倉書店刊,平成24年末刊行開始予定）として刊行する準備を進めている。
- (4) 構築を終えた『現代日本語書き言葉均衡コーパス』は以下の3通りの方法で公開する。
 - A) オンライン全文検索
 - B) オンライン形態論情報検索
 - C) DVD-ROMによる『現代日本語書き言葉均衡コーパス』全データの公開

A)は既に公開を開始している（<http://www.kotonoha.gr.jp/demo/>）。このサイトは誰でも全く自由に利用することができるが、検索方法は全文検索のみであり、検索結果は最大で500件に制限されている（図3参照）。

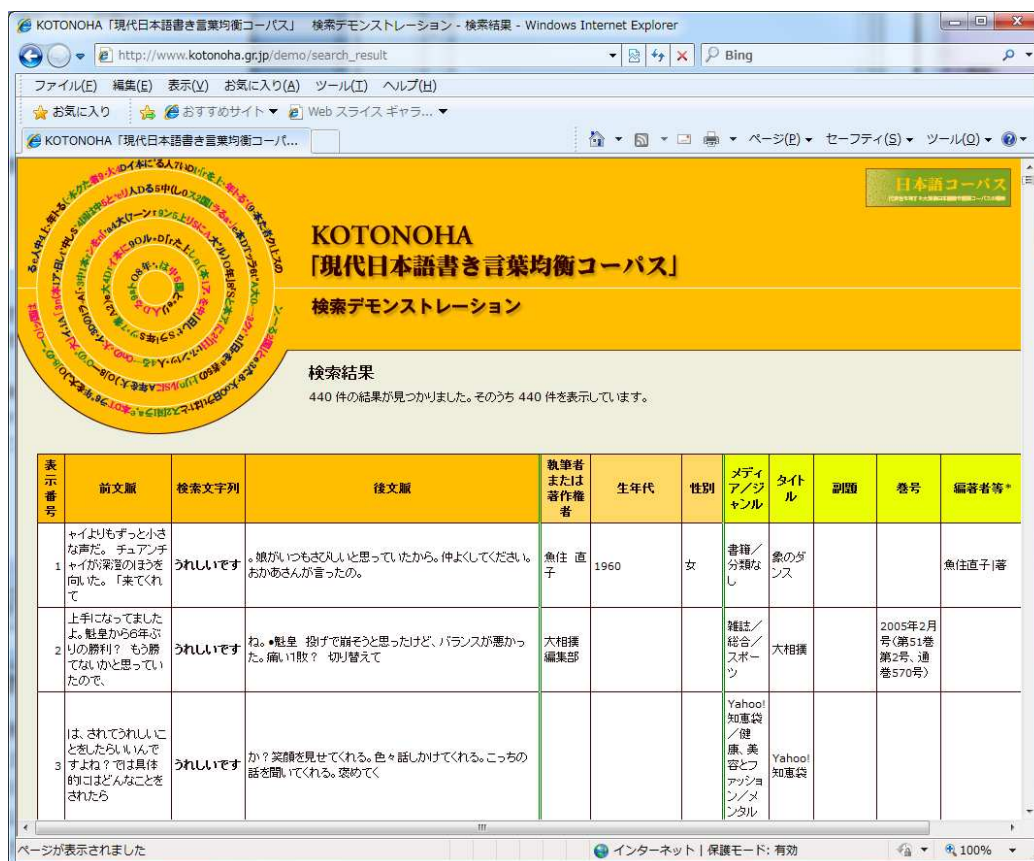


図3: オンライン全文検索サイト

B)は平成23年8月に公開開始の予定である。全文検索に加えて、短単位ないし長単位による形態素解析情報を用いた検索が可能である。利用は当面无償であるが、著作権保護のために事前に利用申請を要求する。検索結果の件数に対する制約はなく、ダウンロードも可能である（図1参照）。

C)は平成23年度中に開始する予定である。利用は有償（コーパス開発環境維持費用として5万円程度を想定）とする。やはり利用申請を要求する。

6. 研究成果の公表の状況

6. 1 主な論文等一覧

4つのカテゴリに分類して代表的な文献を示す。下線は研究分担者、二重下線は代表者、*は corresponding author。 引用回数は判明しているものだけ掲載した。

A) 査読済論文 (192篇、うち71篇は英語で執筆されfull paperで事前査読された国際学会予稿集論文)

- *Yasuharu Den, Nakamura, J., Ogiso, T., & Ogura, H. (2008). A proper approach to Japanese morphological analysis: Dictionary, model, and evaluation. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC2008)*, Marrakech, Morocco, pp. 1019-1024.
- *Ryu Iida, Kentaro Inui, Yuji Matsumoto (2009). "Capturing Saliency with a Trainable Cache Model for Zero-anaphora Resolution," *Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2009)*, pp.647-655. (引用9回)
- *Cao Hongquan, KIKUKO NISHINA (2007). "ERROR ANALYSIS OF JAPANESE ADJECTIVAL COLLOCATIONS FOR AN ERROR DATABASE", *Proceedings CASTEL-J in Hawaii 2007*, pp. 255-258.
- *Chu-Ren Huang, Takenobu Tokunaga, Sohpie Yat Mei Lee. "Asian language processing: current state-of-the-art," *Language Resources and Evaluation*, Vol.40, No.3-4, 203-218, 2006.
- *Ryu Iida, Mamoru Komachi, Kentaro Inui, Yuji Matsumoto (2007). "Annotating a Japanese Text Corpus with Predicate-Argument and Coreference Relations," *ACL Workshop `Linguistic Annotation Workshop*, pp.132-139. (引用20回)
- *Ryu Iida, Kentaro Inui, Yuji Matsumoto (2007). "Zero-anaphora resolution by learning rich syntactic pattern features," *ACM Transactions on Asian Language Information Processing (TALIP)*, Vol 6, Issue 4, Article 12. (引用8回)
- *Mamoru Komachi, Ryu Iida, Kentaro Inui and Yuji Matsumoto (2007). "Learning Based Argument Structure Analysis of Event-nouns in Japanese," *Proceedings of the Conference of the Pacific Association for Computational Linguistics (PACLING)*, pp.120-128. (引用4回)
- *Kikuo Maekawa (2008). "Balanced Corpus of Contemporary Written Japanese." *Proceedings of the 6th Workshop on Asian Language Resources (ALR)*, pp.101-102. (引用6回)
- *Kikuo Maekawa (2010). "Coarticulatory reinterpretation of allophonic variation: Corpus-based analysis of /z/ in spontaneous Japanese." *Journal of Phonetics*, 38(3), pp.360-374.
- *Kikuo Maekawa, Yamazaki, M., Maruyama, T., Yamaguchi, M., Ogura, H., Kashino, W., Ogiso, T., Koiso, H., & Den, Y. (2010). Design, compilation, and preliminary analyses of balanced corpus of contemporary written Japanese. *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC2010)*, Valletta, Malta, pp. 1483-1486.
- *Kyoko Hirose Ohara (2009). "Frame-based contrastive lexical semantics in Japanese FrameNet: The case of risk and kakeru." In Hans C. Boas (Ed.) *Multilingual FrameNets in Computational Lexicography: Methods and Applications*, pp. 163-182. Mouton de Gruyter. (引用14回)
- *Manabu Okumura, Kiyoaki Shirai, Kanako Komiya, Hikaru Yokono (2010). SemEval-2010 Task:Japanese WSD. *The 5th International Workshop on Semantic Evaluation*, pp.69-74. (引用5回)
- *N. Minematsu, Kuroiwa, R., Hirose, K., & Watanabe, M. (2007). "CRF-based statistical learning of Japanese accent sandhi for developing Japanese text-to-speech synthesis systems." *Proceedings of the 6th ISCA Workshop on Speech Synthesis*, Bonn, Germany, pp. 148-153.
- *Vera Sheinman, Neil Rubens, Tokunaga Takenobu (2007). "Commonly Perceived Order within a Category," *Proceedings of OntoLex 2007: from text to knowledge*, ISWC07. (引用2回)
- *Vera Sheinman, Tokunaga Takenobu (2007). "WordSets: Finding Lexically Similar Words for Second Language Acquisition," *Recent Advances in Natural Language Processing*. (引用1回)
- *Kiyoaki Shirai, Hiroshi Ookawa. "Compiling a Lexicon of Cooking Actions for Animation Generation". *21st International Conference on Computational Meeting of the Association for Computational*

- Linguistics*, pp771-778, 2006, (引用3回).
- *Irena Srdanovic Erjavec, *Andrej Bekes, Kikuko Nishina(2007).“ Cluster analysis of suppositional adverbs and clause-final modality.” *Asian and African Studies*, University of Ljubljana Faculty of Arts, vol. XI, no. 3, pp. 21-31.
- *Kazunari Sugiyama, Manabu Okumura (2007), “Personal Name Disambiguation in Web Search Results Based on a Semi-Supervised Clustering Approach”, *Proceedings of the 10th International Conference on Asian Digital Libraries, Lecture Notes in Computer Science (LNCS)*, Springer Verlag. (引用1回).
- *Katsuo Tamaoka & Koizumi, M, (2006). “Issues on the scrambling effects in the processing of Japanese sentences: Reply to Miyamoto and Nakamura (2005) regarding the experimental study by Koizumi and Tamaoka (2004)”. *言語研究*, 129, pp.181-226, 日本言語学会.
- *Katsuo Tamaoka, Matsumoto, M. & Sakamoto, T. (2007). “Identifying empty subjects by modality information: The case of the Japanese sentence-final particles –yo and –ne.” *Journal of East Asian Linguistics*, 16, pp.145-170.
- *Tanaka, J.-I., Tamaoka, K. & Sakai, H. (2007). “Syntactic priming effects on the processing of Japanese sentences with canonical and scrambled word orders.” *認知科学*, 14(2), pp.173-191.
- *Takenobu Tokunaga, Virach Sornlertlamvanich, Thatsanee Charoenporn, Nicoletta Calzolari, Monica Monachini, Claudia Sonia, Chu-Ren Huang, Xia YingJu, Xia YingJu, Yu Hao, Laurent Prevot, Shirai Kiyooki (2006). “Infrastructure for standardization of Asian language resources,” *Proceedings of the COLING/ACL 2006*, pp.827-834. (引用26回)
- *Kiyotaka Uchimoto, & Isahara, H. (2007). “Morphological annotation of a large spontaneous speech corpus in Japanese.” *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI2007)*, Hyderabad, India, pp. 1731-1737.
- *Yotaro Watanabe, Masayuki Asahara, Yuji Matsumoto (2007). “A Graph-Based Approach to Named Entity Categorization in Wikipedia Using Conditional Random Fields”, *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp.649-657. (引用19回)
- *李在鎬・鈴木幸平・永田由香・黒田航・井佐原均(2007).「動詞「流れる」の語形と意味の問題をめぐって」, 『計量国語学』第26巻2号, pp.64-74.
- *飯田龍・小町守・井之上直也・乾健太郎・松本裕治(2010).「述語項構造と照応関係のアノテーション: NAISTテキストコーパス構築の経験から」 *自然言語処理*, Vol.17, No.2, pp.25-50.
- *岩立将和・浅原正幸・松本裕治(2008).「トーナメントモデルを用いた日本語係り受け解析」 *自然言語処理*, Vol.15, No.5, pp.169-185.
- *スルダノヴィッチ イレーナ・ホドスチェエク ボル・ベケシュ アンドレイ・仁科喜久子. (2009) 「ウェブコーパスと検索システムを利用した推量副詞とモダリティ形式の遠隔共起抽出と日本語教育への応用」, *自然言語処理*, Vol. 16, No.4, pp. 29-46.
- *荻野綱男(2006).「形容動詞連体形における「な／の」選択について—田野村氏の結果を WWW で調べる—」 *計量国語学* Vol.25, No.7, pp.309-318.
- *小椋秀樹・相澤正夫(2007).「現代雑誌70誌における漢字の使用実態と常用漢字表—国語施策へのコーパス活用に向けた基礎調査—」 *日本語科学*, 22, pp.125-146.
- *河内昭浩(2010).「作文指導におけるコーパスの活用—高等学校での小論文指導を通して—」 『*解釈*』 56(5・6), pp.27-36, 解釈学会.
- *佐野大樹(2010).「『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」 *機能言語研究*, 5, pp.89-102.
- *柴崎秀子・玉岡賀津雄 (2010). 国語教科書のテキストを基にした小・中学校の学年基準判定式の構築. *日本教育工学会論文誌*. 33(4), pp.449-458.
- *砂川有里子(2007).「分裂文の文法と機能」 『*日本語文法*』 7(2), pp.20-36.
- *曹大峰(2007).「多言語コーパスと日本語研究」 『*日本語科学*』 22, pp.59-78.
- *滝沢直宏(2006).「コーパス利用のためのコンピュータ・リテラシー」 『*日本語教育*』 130, pp.22-31.
- *棚橋尚子(2010).「教科に特徴的な漢字に関する考察—他教科における漢字指導の可能性—」, 『*国語科教育*』 67, pp.11-18, 全国大学国語教育学会.

- *田野村忠温(2008)。「日本語研究の観点からのサーチエンジンの比較評価—Yahoo!とGoogleの比較を中心に—」『計量国語学』第26巻第5号, pp.147-157.
- *田野村忠温(2009)。「サ変動詞の活用のゆれについて・続—大規模な電子資料の利用による分析の精密化—」『日本語科学』第25号, pp.91-103.
- *田野村忠温(2010)。「日本語コーパスとコロケーション—辞書記述への応用の可能性—」『言語研究』第138号, pp.1-23.
- *玉岡賀津雄・木山幸子・宮岡弥生 (印刷中). 新聞と小説のコーパスにおけるオノマトペと動詞の共起パターン. 言語研究, 139.
- *玉岡賀津雄 (2006)。「決定木」分析によるコーパス研究の可能性: 副詞と共起する接続詞「から」「ので」「のに」の文中文末表現を例に. 自然言語処理13(2), pp.169-179.
- *伝康晴・小木曾智信・小椋秀樹・山田篤・峯松信明・内元清貴・小磯花絵 (2007). コーパス日本語学のための言語資源: 形態素解析用電子化辞書の開発とその応用. 日本語科学, 22, pp. 101-122. (引用11回).
- *橋本泰一・吉田恭祐・野口正樹・徳永健伸・田中穂積(2007). 「関係データベースを用いた構文木付きコーパス検索手法」, 自然言語処理, Vol.14, No.4, 3-22.
- *服部匡(2009)。「『～シテイル』と『～シテオル』—戦後の国会会議録における使用傾向調査—」『計量国語学』第27巻第1号, pp.1-17.
- *服部匡(印刷中)。「程度の側面を持つ名詞とそれを量る形容詞類との共起関係—通時的研究—」『言語研究』第140号, 頁未定.
- *前川喜久雄(2007)。「コーパス日本語学の可能性—大規模均衡コーパスがもたらすもの—」日本語科学, 22, pp.13-28. (引用2回).
- *前川喜久雄(2008)。「KOTONOHA『現代日本語書き言葉均衡コーパス』の開発」日本語の研究, 4(1), pp.82-95. (引用2回)
- *前川喜久雄(2008)。「日本語コーパス開発の現状と展望」英語コーパス研究, 15, pp.3-16.(引用2回).
- *丸山直子(2010)。「助詞「に」を伴う<役割>成分—コーパスに基づく分析—」日本語文法, 10(1), pp. 71-87.
- *山内博之(2006)。「第二言語習得過程における言語転位の研究—日本語学習者による「の」の過剰使用を対象に—」『第二言語としての日本語の習得研究』9, pp.104-114.
- *山本和英・齋藤真実(2008)。「用例利用型による文間接続関係の同定」自然言語処理. 15(3), pp.21-51.

B) 無査読論文 (224篇)

- 石井正彦(2009)。「テレビ放送のマルチメディア・コーパス—映像・音声を利用した計量的言語使用研究の可能性—」『阪大日本語研究』21, 1~20頁, 大阪大学大学院文学研究科日本語学講座.
- 石井正彦(2010)。「日本語コーパス言語学の新展開」『日本言語文化』第16輯, 5~21頁, 韓国日本言語文化学会.
- 石黒 圭・西條美紀・仁科喜久子・松見 法男 (2009) 「作文教育のための語彙研究」について『日本語教育』, 日本語教育学会, No. 140, pp. 1-3
- 奥村 学・白井清昭(2008)。「現代日本語書き言葉均衡コーパスを用いた意味解析 — 語義の自動特定, 新語義の発見 —」言語, Vol.37, No.8, pp.66-73.
- 奥村学・白井清昭 (2009)。「代表性のあるコーパスを利用した日本語意味解析」人工知能学会誌, Vol.24, No.5, pp.673-680.
- 小木曾智信(2009). 形態論情報の自動付与とその問題点. 国文学 解釈と鑑賞, 74 (1), pp. 35-43..
- 小木曾智信(2009). コンピュータの辞書. 新「ことば」シリーズ22「辞書を知る」, pp. 114-117.
- 小木曾智信・小椋秀樹・田中牧郎・近藤明日子・伝康晴(2010). 中古和文を対象とした形態素解析辞書の開発. 情報処理学会研究報告, 2010-CH-85, pp. 49-64.
- 荻野綱男(2009)。「コロケーション辞書」国文学解釈と鑑賞, 74 (1), pp.70-78.
- 荻野綱男(2009)。「WWWをコーパスとしてみたときの間違い」言語, 36 (6), pp.6-7.
- 荻野綱男(2009).検索エンジン Google における「単語」語文, 134, pp. 1-4.
- 小椋秀樹(2008)。「『日本語話し言葉コーパス』の言語単位」『日本語学』27-5, pp.72-81.

- 小椋秀樹(2008).「コーパスのための形態論情報」『国文学解釈と鑑賞』74-1, pp.26-34, 至文堂.
- 小椋秀樹(2011).「漢字の使用実態—表外訓・表外字の使用について」国文学解釈と鑑賞, 76, pp.67-75.
- 小原京子(2008).「日本語主観移動表現のコーパス分析:英語との比較から」『慶應義塾大学日吉紀要 言語・文化・コミュニケーション』40号, pp.107-122.
- 河内昭浩(2009).「作文指導における「教科書コーパス」の活用—マッピングと共起語の援用—」, 『国語教室』90, pp.29-33.
- 河内昭浩(2011).「他教科の教科書を使って作文を書く—コーパスを活用した作文指導」, 『月刊国語教育』30(6), pp.78-81, 東京法令出版
- 黒岩龍・*峯松信明・伝康晴・広瀬啓吉 (2007). 単独ラベラによる大規模アクセントデータベースの構築およびそれを利用した統計的アクセント結合処理の検討. 電子情報通信学会技術研究報告, SP2006-174, pp. 31-36.
- 小林ミナ(2009).「『基本的な文法項目』とは何か」『日本語教育の過去・現在・未来 — 第5巻 文法』 pp.40-61.
- 近藤明日子・田中牧郎(2008).「学校教科書の語彙—語種を観点として—」, 『日本語学』27(10), p26-35.
- 佐野大樹(2010).「ブログにおける評価情報の分類と体系化—アプレイザル理論を用いて」, 信学技報, NLC2009-39, pp.37-42.
- 白井清昭(2009).「コーパスにおける語の意味の自動識別」, 国文学 解釈と鑑賞. Vol. 74, No. 1, pp.61-69, (引用1回).
- 鈴木一史(2007).「教科書コーパスによる国語科学習語彙の選定方法の検討」, 『東京大学教育学部附属論集』51.
- 鈴木一史(2008).「国語教育における『語彙』指導」『日本語学』27(10), pp.16-25.
- 鈴木一史(2009).「コーパスを用いた授業の可能性」『月刊国語教育』356, pp.34-37.
- 砂川有里子(2008).「事象叙述の有題文と無題文—自然会話を用いたケーススタディー—」『叙述類型論』 pp. 115-137.
- 橋田浩一・和泉憲明.「オントロジーに基づく知識の構造化と活用」(2007). 情報処理, Vol.48, No.8, pp.843-848.
- 橋本直幸・山内博之(2008).「日本語教育のための語彙リストの作成」『日本語学』 pp. 50-58.
- 服部匡(2007).「因子分析を用いた程度副詞と述語等の共起関係分析の試み—新聞コーパスのデータから—」『同志社女子大学総合文化研究所紀要』第24号, pp.98-109.
- 服部匡(2007).「大規模コーパスを用いた副詞『全然』の共起特性の調査—朝日新聞とYahoo!知恵袋の比較—」『同志社女子大学学術研究年報』第58号, pp.1-8.
- 棚橋尚子(2007).「漢字学習におけるルビの有効性」, 『奈良教育大学国文—研究と教育—』30, pp.77-89, 奈良教育大学
- 前川喜久雄(2008).「話し言葉と書き言葉」日本語学, 27(5), pp.23-33(引用1回).
- 前川喜久雄(2009).「30年の時間幅において観察される語義およびコロケーションの変化—『現代日本語書き言葉均衡コーパス』の予備的分析—」堀正広・浮網茂信・西村秀夫・小迫勝・前川喜久雄『コロケーションの通時的的研究—英語・日本語研究の新たな試み』ひつじ書房, pp.183-198.
- 前川喜久雄(2009).「KOTONOHA『現代日本語書き言葉均衡コーパス』における著作権処理」漢字文献情報処理研究,10, pp.30-35.
- 前川喜久雄(2009).「代表性を有する大規模日本語書き言葉コーパスの構築」人工知能学会誌, 24(5), pp.616-622.
- 前川喜久雄(2010).「コーパス構築と著作権保護」人工知能学会誌, 25(5), pp.628-632.
- 前川喜久雄 (印刷中)「コーパスと言語学」松本裕治編『言語と情報科学』(シリーズ朝倉「言語の可能性」<6>).
- 松本裕治 (2009).「コーパスへの自動アノテーションツールとアノテーション支援環境の構築」. 人工知能学会誌, Vol.24, No.5, pp.632-639.
- 松本裕治 (2009).「統語情報の付与」国文学と鑑賞, Vol.74, No.1, pp.44-52.
- 松本裕治・大山浩美 (2009).「言語処理による作文支援・語彙学習への可能性について」. 日本語教育「特集 作文教育のための語彙研究」, Vol.140, pp.37-47, 日本語教育学会.

- 丸山岳彦(2008). 「『日本語話し言葉コーパス』の節単位情報」 『日本語学』 27(5), pp.82-89.
- 丸山岳彦(2008). 「日本語コーパスの現状」 『国文学解釈と鑑賞』 74-1, pp.122-130 (引用1回)
- 丸山岳彦(2009). 「作文の文体情報—『現代日本語書き言葉均衡コーパス』から見えるもの—」, 『日本語教育』 140号, 日本語教育学会.
- 田中弥生(2011). 「Yahoo!知恵袋における質問の修辞ユニット分析-脱文脈化—文脈化の程度による分類—」 信学技報, NLC2010-33, pp.13-18.
- 田野村忠温(2008). 「大規模な電子資料に見る現代日本語の動態」 『待兼山論叢』 第42号文化動態論篇, pp.55-76, 大阪大学大学院文学研究科.
- 田野村忠温(2009). 「特集 日本語研究とコーパス」 コーパスと文法研究」 『国文学 解釈と鑑賞』 第74巻第1号, pp.79-87.
- 田野村忠温(2009). 「コーパスからのコロケーション情報抽出—分析手法の検討とコロケーション辞典項目の試作—」 『阪大日本語研究』 21, pp.21-41, 大阪大学大学院文学研究科日本語学講座.
- 田野村忠温(2009). 「コーパスを用いた日本語研究の精密化と新しい研究領域・手法の開発」 『人工知能学会誌』 第24巻第5号, pp. 647-655.
- 伝康晴(2009). 多様な目的に適した形態素解析システム用電子化辞書. 人工知能学会誌, 24, pp. 640-646.
- 服部匡 (印刷中). 「言語資料としての国会会議録の特徴(1)—本会議と委員会等との比較—」 『同志社女子大学日本語日本文学』 23号.
- 矢澤真人(2007). 「ユビキタス辞書の時代」 日本語学, pp.26-8, pp.58-66.
- 山崎誠 (2007). 「国立国語研究所の語彙調査の歴史と課題」 Sokutei Report VOL.6 (東京大学大学院教育学研究科教育研究創発機構教育測定・カリキュラム開発(ベネッセコーポレーション) 講座), pp.168-186.
- 山崎誠 (2007). 「国立国語研究所の言語コーパス整備計画「KOTONOHA」の紹介」 『漢字文献情報処理研究』 8, pp.180-183.
- 山崎誠 (2008). 「国立国語研究所における諸研究—語彙調査の系譜の中心にして—」, 『国文学解釈と鑑賞』 74-1, pp.183-191.
- 山崎誠(2009). 「代表性を有する大規模日本語書き言葉コーパスの構築」 人工知能学会誌, 24(5), pp.623-631.

C) 著書 (10篇)

- Kyoko Hirose Ohara & Nikiforidou, Kiki* (Eds.) (2010). *Constructions and Frames*. John Benjamins Publishing Company.
- Yukio Tono (2007). *Spoken Corpora in Applied Linguistics*. Peter Lang Pub Inc, 264p.
- 柴崎秀子(2006). 『第二言語テキスト理解と読み手の知識』 風間書房.
- 滝沢直宏(2006). 『コーパスで一目瞭然』 小学館, 207p.
- 石井正彦(2007). 『現代日本語の複合語形成論』 ひつじ書房.
- 投野由紀夫(2007). 『日本人中高生 1 万人の英語コーパス JEFLL Corpus —中高生が書く英文の実態とその分析—』 小学館, 191p.
- 山内博之編(2008). 『日本語教育スタンダード試案 語彙』 ひつじ書房, 117p.
- 山内博之編(2009). 『プロフィシェンシーから見た日本語教育文法』 ひつじ書房, 187p.
- 投野由紀夫(監修) & Christian James(英文校閲)(2010). 『フェイバリット英単語・熟語<テーマ別>コーパス1800』 東京書籍, 272p.
- 砂川有里子他編(2010). 『日本語教育研究への招待』 くろしお出版, 300p.

D) 口頭発表 (541篇)

- Kikuo Maekawa. "Design of a Balanced Corpus of Contemporary Written Japanese." Proceedings of Symposium on Large-Scale Knowledge Resources (LKR2007), pp.55-58, 2007:03.
- Kikuo Maekawa. "KOTONOHA and BCCWJ: Development of a Balanced Corpus of Contemporary Written Japanese." Corpora and Language Research: Proceedings of the First International Conference on Korean Language, Literature, and Culture. Seoul, Yonsei University, pp.158-177, 2007:02,引用2件.

- Kikuo Maekawa. "Compilation of the Balanced Corpus of Contemporary Written Japanese in the KOTONOHA Initiative." Proc. Second International Symposium on Universal Communication, pp.169-172, 2008:12.
- Kikuo Maekawa. "KOTONOHA: A Corpus Compilation Initiative at the National Institute for Japanese Language," Keynote speech at the 22nd International Conference on the Computer Processing of Oriental Languages, Hong Kong Polytechnic University, Hong Kong, 2009:03.
- 相澤正夫・小椋秀樹(2008).「白書コーパスに基づく常用漢字の使用実態調査」第21回社会言語科学会研究大会, 2008年3月22日, 東京女子大学.
- 秋元祐哉・丸山岳彦・吉田谷幸宏・山崎誠・柏野和佳子・稲益佐知子・前川喜久雄(2007).「書き言葉の総量を捉える—書き言葉はどれだけ生産されるのか—」言語処理学会第13回年次大会 [NLP2007]予稿集pp.708-711, 2007年3月21日, 龍谷大学.
- 飯田龍・小町守・乾健太郎・松本裕治.「NAISTテキストコーパス: 述語項構造と共参照関係のアノテーション」, 情報処理学会第177回自然言語処理研究会, 2007-NL-177, pp.71-78, 2007.
- 飯田龍・小町守・乾健太郎・松本裕治.「日本語書き言葉を対象とした述語項構造と共参照関係のアノテーション: NAISTテキストコーパス開発の経験から」, 言語処理学会第13回年次大会論文集, 2007.
- 石井正彦「外来語の20世紀」日本語学会平成19年度春季大会、関西大学、2007年5月26日
- スルダノヴィッチエリヤヴェッシ イレーナ・*ベケシュアンドレイ・仁科喜久子 (2008) 「Sketch Engineを利用した遠隔共起の抽出と日本語教育への応用」, 言語処理学会第14回年次大会ワークショップ「教育・学習を支援する言語処理」論文集, 言語処理学会, pp. 39-42
- 小木曾智信・小椋秀樹・伝康晴 (2007). 日本語研究に適した形態素解析ソフトウェア—「unidic」と「茶まめ」—. 日本語学会2007年度秋季大会予稿集, pp. 255-262.
- 小木曾智信・伝康晴・渡部涼子・近藤明日子 (2009). 現代語コーパスの利用による近代語形態素解析の精度向上. 言語処理学会第15回年次大会発表論文集, pp. 801-804.
- 小椋秀樹・小木曾智信・小磯花絵・富士池優美・相馬さつき(2007).「現代日本語書き言葉均衡コーパス」の短単位解析について」言語処理学会第13回年次大会[NLP2007]予稿集pp.720-723, 2007年3月21日, 龍谷大学.
- 小椋秀樹・小木曾智信・原裕・小磯花絵・富士池優美(2008).「形態素解析用辞書UniDicへの語種情報の実装と政府刊行白書の語種比率の分析」言語処理学会第14回年次大会[NLP2008]予稿集, pp.935-938, 2008年3月20日, 東京大学.
- 小原京子(2008).「コーパスに基づく日本語主観移動表現のフレーム意味論的分析: 英語との比較から」日本認知科学会第25回大会.
- 柏野和佳子(2006).「国語辞典における多義語の意味区分の比較」第23回ことば工学研究会資料集 SIG-SE-A601-4 pp.37-43, 2006年8月5日, 神奈川大学.
- 柏野和佳子(2007).「国語辞典における多義語の意味記述の比較」言語処理学会第13回年次大会 [NLP2007]予稿集pp.863-866, 2007年3月21日, 龍谷大学.
- 河内昭浩 (2010) 「語彙を豊かにする作文指導—「コーパス」を活用した実践と今後の教材開発に向けて—」, 『全国大学国語教育学会 国語教育研究 第119回鳴門大会研究発表要旨集』, pp.186-189, 全国大学国語教育学会
- 北村雅則(2007).「<驚き・感慨>を表すモノダ文の解釈と構造」日本語学会2007年度春季大会予稿集, pp.89-96, 2007年5月27日, 関西大学.
- 黒岩龍・*峯松信明・伝康晴・広瀬啓吉 (2007). 大規模アクセントラベリングコーパスの構築とそれに基づくハイブリッド型アクセント結合処理. 言語処理学会第13回年次大会発表論文集, pp. 724-727. 不詳.
- 小磯花絵・小木曾智信・小椋秀樹・富士池優美・宮内佐夜香(2008).「『現代日本語書き言葉均衡コーパス』にもとづくジャンル間の文体差に関わる要因の分析」第22回社会言語科学会研究大会, 2008年9月13日, 愛知大学.
- 小磯花絵・小木曾智信・小椋秀樹・富士池優美・宮内佐夜香 (2008).『現代日本語書き言葉均衡コーパス』にもとづくジャンル間の文体差に関わる要因の分析. 社会言語科学会第22回研究大会発表論文集, pp. 192-195.

- 小林ミナ(2006). 「記述文法と規範文法のインターフェイスとしての『教育文法』」早稲田大学日本語教育学会, 2006.9.9, 早稲田大学.
- 近藤明日子 (2008) 「中学校教科書の教科別特徴語の抽出—理科を例として—」, 『特定領域研究「日本語コーパス」平成19年度公開ワークショップ(研究成果報告会)予稿集』, pp.181-186
- 斎藤博昭・小原京子・曾根孝明・久保谷俊太 (2008). 「日本語コーパスを用いた日本語フレームネットのアノテーション環境」, 特定領域研究「日本語コーパス」平成19年度公開ワークショップ.
- 佐野大樹・丸山岳彦(2008). 「システミック文法に基づく書きことばの複雑さ測定—日本語大規模コーパスを用いた語彙密度計測—」言語処理学会第14回年次大会[NLP2008]予稿集, pp.1097-1100, 2008年3月20日, 東京大学.
- 佐野大樹(2008). 「大規模バランストコーパスにおけるテキスト分類に向けて—語彙密計測からみたコンテキスト情報—」日本機能言語学会第16回秋季大会, 2008年10月12日, お茶の水女子大学.
- 白井清昭・徳永健伸. 「呼応する名詞の包含関係に着目した助数詞オントロジーの自動構築と評価」, 情報処理学会 自然言語処理研究会. Vol.NL-181. pp.127-134, 2007
- 山崎誠・丸山岳彦・柏野和佳子・前川喜久雄・稲益佐知子・秋元祐哉・吉田谷幸宏(2006). 「現代日本語書き言葉均衡コーパスのサンプリング方法について」計量国語学会第50回大会, 2006年9月30日, 国立国語研究所.
- 砂川有里子(2008). 「日本語教育におけるコーパス活用の可能性」日本語教育世界大会2008, 2008.7.13, 釜山外国語大.
- 曹 紅荃・仁科喜久子 (2008) 「日本語学習者が産出した形容詞名詞共起表現の誤用データベースの構築」, 言語処理学会第14回年次大会ワークショップ「教育・学習を支援する言語処理」論文集, 言語処理学会, pp. 27-30
- 高田智和・間淵洋子・西部みちる・北村雅則・山口昌也(2007). 「文字コードとタグによる漢字字体の記述」言語処理学会第13回年次大会[NLP2007]予稿集pp.712-715, 2007年3月21日, 龍谷大学.
- 竹内孔一・*下村拓也 (2008). 動詞語義を推定するための語義付与コーパスの作成: 言語処理学会第14回年次大会, PD1-7, pp.273-276.
- 滝沢直宏(2007). 「コーパスを用いた英語研究の方法」日本言語学会, 2007.6.17, 麗澤大学.
- 田中牧郎・金愛蘭・桐生りか・近藤明日子 (2008) 「コーパスによる難解語・重要語の抽出—医療用語を例に—」, 『社会言語科学会第21回大会発表論文集』, pp.296-299, 社会言語科学会
- 田中弥生(2008). 「クチコミサイトにおける世代別・媒体別言語表現の分析」言語処理学会第14回年次大会[NLP2008]予稿集, pp.911-914, 2008年3月20日, 東京大学.
- 谷口雄作・新保仁・浅原正幸・松本裕治. 「タグ付きコーパス検索ツール『茶杓』」, 言語処理学会第13回年次大会論文集, 2007.
- 千葉庄寿(2008). 「コリゲーションの抽出における形態統語情報の役割」言語処理学会第12回年次大会, 2008.3.20, 東京大学. 服部匡「因子分析を用いた程度副詞と述語等の共起関係の研究試論」平成18年度公開ワークショップ、時事通信ホール、2007年3月18日
- 伝康晴・中村純平・小木曾智信・小椋秀樹(2008). 「語種情報を用いた同表記異音語の解消」言語処理学会第14回年次大会[NLP2008]予稿集, pp.69-72, 2008年3月18日, 東京大学.
- 服部匡「気づかれにくい語結合の傾向とその変化—共時的/通時的コーパスのデータから—」第26回表現学会近畿例会、同志社大学、2010年2月27日
- 服部匡「国会会議録データと現代語の通時変化」日本語学会2010年度秋季大会ワークショップ「コーパス日本語学の新展開—コーパスと方法論の多様化—」、愛知大学、2010年10月23日
- 藤井聖子・上垣渉(2008)「支援動詞構文における事態性名詞と動詞との項共有と連結性: 『日本語コーパス』を用いた分析」日本言語学会第136回大会.
- 富士池優美・*小椋秀樹・小木曾智信・小磯花絵・内元清貴・相馬さつき・中村壮範 (2008). 「現代日本語書き言葉均衡コーパス」の長単位認定基準について. 言語処理学会第14回年次大会発表論文集, pp. 931-934. 2008年3月20日, 東京大学.
- 前川喜久雄「大規模言語資源の開発とその問題点(特に著作権処理について)」WebDB Forum 2008 特別セッション「企業の巨大データ徹底解剖 - 新たな研究の可能性と産学連携 -」2008:12.
- 前川喜久雄「国立国語研究所のコーパス整備計画KOTONOHA—その現状と問題点—」電子情報通信学会技術研究報告思考と言語, TL2008-1, pp.82-95, 2008:05.

- 前川喜久雄「国立国語研究所における言語資源」2009言語資源シンポジウム「言語・音声データの学術利用に向けて」(国立情報学研究所),p.90-93,2009:10.
- 丸山岳彦・柏野和佳子・稲益佐知子・秋元祐哉・吉田谷幸宏・山崎誠(2007).「書き言葉の構造を捉える—書き言葉の多様な構造とサンプリング手法—」言語処理学会第13回年次大会[NLP2007]予稿集pp.704-707, 2007年3月21日, 龍谷大学.
- 丸山岳彦・相馬さつき(2007).『日本語話し言葉コーパス』日本言語学会第134回大会公開シンポジウム関連デモンストレーション, 2007年6月17日, 麗澤大学.
- 宮内佐夜香(2009).「丁寧対文内における従属句の文体と接続助詞について—『現代日本語書き言葉均衡コーパス』を資料として—」, 日本語学会2009年度秋季大会, 2009年11月1日, 島根大学.
- 森本祥子・前川喜久雄・小沼悦・新井田貴之・長谷川愛・大石有香・神野博子・竹内ゆかり(2008).「『現代日本語書き言葉均衡コーパス』における著作権処理: 2年間の経験から」特定領域研究「日本語コーパス」平成19年度公開ワークショップ(研究成果発表会)予稿集, pp.47-50, 2008年3月15日, 時事通信ホール.
- 山口昌也・小木曾智信(2007).「言語資料の共有, 利用を支援する環境の実現」日本語学会2007年度秋季大会予稿集, pp.247-254, 2007年11月18日, 沖縄国際大学.
- 山崎誠・丸山岳彦・山口昌也・小椋秀樹・森本祥子・柏野和佳子・佐野大樹・高田智和・間淵洋子・北村雅則・小木曾智信・小磯花絵・富士池優美・小沼悦・田中牧郎・前川喜久雄(2007).「現代日本語書き言葉均衡コーパスの設計と検索デモンストレーション」日本語学会2007年度秋季大会(沖縄国際大学)予稿集, pp.239-246, 2007年11月18日, 沖縄国際大学.
- 山田篤*(2007). 数字列への読み付与—NumTransとChaOne— 特定領域研究「日本語コーパス」平成19年度全体会議予稿集, pp.85-90.
- 吉橋健治・仁科喜久子(2008)「形態素解析に依存しない日本語係り受け解析」, 言語処理学会第14回年次大会発表論文集, 言語処理学会, pp. 81-82

E) 特定領域で刊行した報告書 (52冊 12,178頁)

総括班

- 総括班(2006).『特定領域「日本語コーパス」平成18年度全体会議予稿集』, 52p.
- 総括班(2007).『特定領域研究「日本語コーパス」平成18年度公開ワークショップ(研究成果報告会)予稿集』, 186p.
- 総括班(2007).『特定領域「日本語コーパス」平成19年度全体会議予稿集』, 120p.
- 総括班(2008).『特定領域研究「日本語コーパス」平成19年度公開ワークショップ(研究成果報告会)予稿集』, 230p.
- 総括班(2008).『特定領域「日本語コーパス」平成20年度全体会議予稿集』, 138p.
- 総括班(2009).『特定領域研究「日本語コーパス」平成20年度公開ワークショップ サテライトセッション予稿集』, 150p.
- 総括班(2009).『特定領域研究「日本語コーパス」平成20年度公開ワークショップ(研究成果報告会)予稿集』, 245p.
- 総括班(2009).『特定領域「日本語コーパス」平成21年度全体会議予稿集』, 168p.
- 総括班(2010).『特定領域研究「日本語コーパス」平成21年度公開ワークショップ サテライトセッション予稿集』, 118p.
- 総括班(2010).『特定領域研究「日本語コーパス」平成21年度公開ワークショップ(研究成果報告会)予稿集』, 264p.
- 総括班(2010).『特定領域「日本語コーパス」平成22年度全体会議予稿集』, 210p.
- 総括班(2011).『特定領域研究「日本語コーパス」平成22年度公開ワークショップ(研究成果報告会)予稿集』, 552p.
- 総括班(2011).『特定領域研究「日本語コーパス」研究成果報告』(DVD).

データ班

- 小椋秀樹(2007).『『現代日本語書き言葉均衡コーパス』短単位規程集 Version 1.2』, 140p.
- 丸山岳彦・秋元祐哉(2007).『『現代日本語書き言葉均衡コーパス』におけるサンプル構成比の算出法—現代日本語書き言葉の文字数調査—』, 76p.

- 丸山岳彦・秋元祐哉(2008). 『『現代日本語書き言葉均衡コーパス』におけるサンプル構成比の算出法(2)ーコーパスの設計とサンプルの無作為抽出法ー』, 52p.
- 柏野和佳子・丸山岳彦・秋元祐哉・稲益佐知子・佐野大樹・田中弥生・山崎誠(2008). 『『現代日本語書き言葉均衡コーパス』における書籍サンプルの多様性』, 36p.
- 山口昌也・高田智和・北村雅則・間淵洋子・小林正行・西部みちる(2008). 『『現代日本語書き言葉均衡コーパス』における電子化フォーマット ver.2.0』, 162p.
- 小椋秀樹・小磯花絵・富士池優美・原裕(2008). 『『現代日本語書き言葉均衡コーパス』形態論情報規程集』, 200p.
- 柏野和佳子・丸山岳彦・稲益佐知子・田中弥生・秋元祐哉・佐野大樹・大矢内夢子・山崎誠(2009). 『『現代日本語書き言葉均衡コーパス』における収録テキストの抽出手順と事例』, 104p.
- 佐野大樹・丸山岳彦・山崎誠・柏野和佳子・秋元祐哉・稲益佐知子・田中弥生・大矢内夢子(2009). 『語彙密度を利用した『現代日本語書き言葉均衡コーパス』テキスト分類の試み』, 62p.
- 小椋秀樹・小磯花絵・富士池優美・原裕(2009). 『『現代日本語書き言葉均衡コーパス』形態論情報規程集改定版』, 250p.
- 高田智和・小林正行・間淵洋子・大島一・西部みちる・山口昌也(2009). 『JIS X 0213:2004運用の検証』, 260p.
- 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・原裕(2010). 『『現代日本語書き言葉均衡コーパス』形態論情報規程集第3版』, 298p.
- 丸山岳彦・山崎誠・柏野和佳子・佐野大樹・秋元祐哉・稲益佐知子・田中弥生・大矢内夢子(2011). 『『現代日本語書き言葉均衡コーパス』におけるサンプリングの原理と運用』, 92p.
- 丸山岳彦・山崎誠・柏野和佳子・佐野大樹・秋元祐哉・稲益佐知子・田中弥生・大矢内夢子(2011). 『『現代日本語書き言葉均衡コーパス』に含まれるサンプルの設計と実装』, 154p.
- 西部みちる・大島一・間淵洋子・小林正行・田島孝治・高田智和・山口昌也(2011). 『『現代日本語書き言葉均衡コーパス』における電子化テキストの構築』, 396p.
- 山口昌也・高田智和・北村雅則・間淵洋子・大島一・小林正行・西部みちる(2011). 『『現代日本語書き言葉均衡コーパス』における電子化フォーマット ver2.2』, 166p.
- 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕(2011). 『「現代日本語書き言葉均衡コーパス」形態論情報規定集第4版(上)』 「日本語コーパス」データ班, 132p.
- 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕(2011). 『「現代日本語書き言葉均衡コーパス」形態論情報規定集第4版(下)』 「日本語コーパス」データ班, 228p.
- 電子化辞書班**
- 小木曾智信・中村壮範(2009). 『現代日本語書き言葉均衡コーパス』形態論情報データベースの設計と実装』, 141p.
- 小木曾智信・中村壮範(2011). 『「現代日本語書き言葉均衡コーパス」形態論情報データベースの設計と実装 改訂版』, 145p.
- 日本語学班**
- 田野村忠温・服部匡・杉本武・石井正彦(2007). 『コーパスを用いた日本語研究の精密化と新しい研究領域・手法の開発Ⅰ』 特定領域研究「日本語コーパス」日本語学班平成18年度研究成果報告書, 165p.
- 田野村忠温・服部匡・杉本武・石井正彦(2007). 『コーパス日本語学ガイドブック』、特定領域研究「日本語コーパス」日本語学班, 199p.
- 田野村忠温・服部匡・杉本武・石井正彦(2008). 『コーパスを用いた日本語研究の精密化と新しい研究領域・手法の開発Ⅱ』 特定領域研究「日本語コーパス」日本語学班平成19年度研究成果報告書, 207p.
- 田野村忠温・服部匡・杉本武・石井正彦(2009). 『コーパスを用いた日本語研究の精密化と新しい研究領域・手法の開発Ⅲ』 特定領域研究「日本語コーパス」日本語学班平成20年度研究成果報告書, 211p.
- 田野村忠温・服部匡・杉本武・石井正彦(2010). 『コーパスを用いた日本語研究の精密化と新しい研究領域・手法の開発Ⅳ』 特定領域研究「日本語コーパス」日本語学班平成21年度研究成果報告書, 182p.

田野村忠温・服部匡・杉本武・石井正彦(2010).『コーパス日本語学の新展開』特定領域研究「日本語コーパス」日本語学班平成18年度～平成22年度研究成果報告書, 500p.

日本語教育班

砂川有里子・井上優・小林ミナ・滝沢直宏・投野由紀夫・山内博之 (2008).『代表性を有する書き言葉コーパスを利用した日本語教育研究』文部科学省科学研究費補助金特定領域研究「代表性を有する大規模日本語書き言葉コーパスの構築：21世紀の日本語研究の基盤整備」日本語教育班平成19年度研究成果報告書, 269p.

砂川有里子・井上優・小林ミナ・滝沢直宏・投野由紀夫・山内博之 (2010).『代表性を有する書き言葉コーパスを利用した日本語教育研究』文部科学省科学研究費補助金特定領域研究「代表性を有する大規模日本語書き言葉コーパスの構築：21世紀の日本語研究の基盤整備」日本語教育班平成20-21年度研究成果報告書, 318p.

言語政策班

田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・小椋秀樹・鈴木一史・近藤明日子・平山允子・金愛蘭・桐生りか (2008)『特定領域研究「日本語コーパス」言語政策班中間報告書 言語政策に役立つ, コーパスを用いた語彙表・漢字表等の作成と活用』, 291p.

田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・鈴木一史・近藤明日子・平山允子 (2009)『「教科書コーパス」中学校教科書語彙表(短単位)(1)』, 540p.

田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・鈴木一史・近藤明日子・平山允子 (2009)『「教科書コーパス」中学校教科書語彙表(短単位)(2)』, 539p.

田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・鈴木一史・近藤明日子・平山允子 (2009)『「教科書コーパス」中学校教科書語彙表(疑似長単位)(1)』, 664p.

田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・鈴木一史・近藤明日子・平山允子 (2009)『「教科書コーパス」中学校教科書語彙表(疑似長単位)(2)』, 661p.

田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・鈴木一史・近藤明日子・平山允子 (2009)『「教科書コーパス」中学校教科書国語科特徴語彙表』, 508p.

田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・鈴木一史・近藤明日子・平山允子 (2009)『「教科書コーパス」中学校教科書社会科特徴語彙表』, 503p.

田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・近藤明日子・河内昭浩・鈴木一史・平山允子(2011)『言語政策に役立つ, コーパスを用いた語彙表・漢字表等の作成と活用 (特定領域研究「日本語コーパス」言語政策班報告書)』, 254p.

辞書編集班

荻野綱男 (編) (2007)『コーパスを利用した国語辞典編集法の研究』(特定領域研究「日本語コーパス」辞書編集班) 204p.

荻野綱男 (編) (2008)『コーパスを利用した国語辞典編集法の研究』(特定領域研究「日本語コーパス」辞書編集班) 151p.

荻野綱男 (編) (2009)『コーパスを利用した国語辞典編集法の研究』(特定領域研究「日本語コーパス」辞書編集班) 163p.

荻野綱男 (編) (2010)『コーパスを利用した国語辞典編集法の研究』(特定領域研究「日本語コーパス」辞書編集班) 122p.

6. 2 ホームページ

A) 特定領域研究「日本語コーパス」ホームページ <http://www.tokuteicorpus.jp/>

一般広報用に利用。アクセス数不詳。

B) 『現代日本語書き言葉均衡コーパス』検索デモサイト <http://www.kotonoha.gr.jp/demo/>

『現代日本語書き言葉均衡コーパス』データの全文検索が可能。通算アクセス数14万件以上。

C) UniDicホームページ <http://www.tokuteicorpus.jp/dist/>

本領域で開発した形態素解析用辞書UniDicのホームページ。登録ユーザー数3,755名。

D) 「日本語コーパス」領域専用ホームページ <http://tokuteicorpus.jp/nucleus/index.php>

領域内での情報・データ共有に利用。

6. 3 公開発表

A) 公開研究成果発表会（公開ワークショップ）

以下のとおり各年度末に開催した。参加人数は受付で記帳した人数。本領域の関係者80~90名を差し引いた数が領域外からの参加者数である。

1. 平成18年度（2007年3月17~18日 時事通信ホール）発表27件 参加人数228名
2. 平成19年度（2008年3月15~16日 時事通信ホール）発表44件 参加人数227名
3. 平成20年度（2009年3月15~16日 東京工業大学）発表58件 参加人数292名
4. 平成21年度（2010年3月15~16日 東京工業大学）発表56件 参加人数230名
5. 平成22年度（2011年3月14~16日 時事通信ホール）発表78件 参加申込299名（震災のため中止）
6. 『現代日本語書き言葉均衡コーパス』完成記念講演会（2011年8月2,3日開催予定 JA共済会館）

B) 各研究班が独自に開催した講習会・研究会

データ班：

1. 『現代日本語書き言葉均衡コーパス』仕様説明会
開催日時：2006年11月7日 開催場所：NAIST東京事務所・5階 508リエゾンルーム
内容：構築するコーパスのサンプリング方法及び形態論情報の説明と領域内に公開する白書データの解説を行った。参加者49名。
2. 特定領域研究「日本語コーパス」2007年度データ説明会
開催日時：2007年11月5日 開催場所：埼玉大学東京ステーションカレッジ
内容：領域内に公開するデータについて、サンプリング方法、電子化形式を説明し、短単位による解析データ、全文検索システム「ひまわり」の紹介を行った。参加者52名。

ツール班：

3. 科学技術振興調整費新興分野人材養成プログラム「自然言語処理技術」セミナー。
開催日時：2007年9月4日（月）開催場所：京都大学学術情報メディアセンター（京都）
内容：「『茶釜』/『南瓜』を用いた形態素解析・係り受け解析」，および「『茶器』によるコーパス管理・検索」というタイトルでこれらのツールに関する講習を行った。参加者50名
4. ツール講習会
開催日時：2007年11月25日（日）
開催場所：キャンパスイノベーションセンター5F，リエゾンコーナー501（東京都田町）
内容：形態素解析・係り受け解析コーパス管理ツール「茶器」，汎用アノテーションツールSLAT，および，Webベースのコーパス検索ツール「茶杓」に関する講習会を，特定領域研究のメンバーに対して行った。参加者35名。
5. 拡大ツール班会議
開催日時：2008年8月28日、開催場所：キャンパスイノベーションセンター（東京都田町）
内容：ツール班の活動と進捗を領域内メンバーに紹介し，今後の活動について意見交換を行った。茶器へのコーパスのポート，エラーメッセージ，マニュアルの整備，コーパス中の単語の一覧表示，種々の用例抽出の支援機能など，様々な要望が寄せられ，今後の開発の参考とした。参加者25名。
6. 自然言語処理技術講習会（京都大学主催）
開催日時：2008年9月8日9日、開催場所：京都大学学術情報メディアセンター（京都）
内容：茶器のチュートリアルと実習を行った。講習会へ向けて，茶器の簡易インストラを作成した。インストラの指示に従うことにより，コーパスの解析に必要な言語解析ツールやデータベースシステムを簡単にインストールできるようになった。参加者約60名。
7. ツール講習会
開催日時：2008年11月28日、開催場所：キャンパスイノベーションセンター

内容：主に領域内メンバーを対象として、ツール班で開発しているツールの講習会を行う。茶器については、データベースをネットワーク経由でアクセスできるようになったので、奈良先端大のサーバにあるコーパスを検索対象にして講習を行った。また、本年度に一般公開されたモニター公開コーパスを茶器にインストールするためのDVDの配布、および、モニター公開コーパス登録者に対してNAISTの茶器データベースにアクセスするためのユーザー名とパスワードを発行した。参加者約20名。

8. 自然言語処理技術講習会（京都大学主催）

開催日時：2009年9月30日、開催場所：京都大学学術情報メディアセンター（京都）

内容：茶器のチュートリアルと実習を行った。茶器の簡易インストーラを作成し、コーパスの解析に必要な言語解析ツールやデータベースシステムを簡単にインストールできるようにした。また、未解析のコーパスファイルに対して、形態素解析器MeCabと係り受け解析器CaboChaを実行するための簡易フォーマットを作成した。参加者約50名。

9. ツール講習会

開催日時：2009年11月9日、開催場所：奈良先端科学技術大学院大学情報科学研究科

内容：主に領域内メンバーを対象として、ツール班で開発しているツール類の講習会を行った。茶器については、日本語コーパス外部公開版に含まれる形態素情報付きのコアデータのSQLiteファイルを配布し、茶器の基本機能に関する講習を行った。参加者約20名。

10. ツール講習会

開催日時：2010年11月20日、開催場所：キャンパスイノベーションセンター多目的室1（東京都田町）

内容：コーパス管理ツール茶器の現状報告と使用方法に関するチュートリアルを行った。参加者15名。

日本語教育班：

11. ワークショップ「日本語教育とコーパス」

開催日時：2010年6月12日、開催場所：麗澤大学言語研究センター
発表5件、参加者約30名。

12. ワークショップ「エディターを使ったコーパス検索」

開催日時：2010年7月8日～9日、開催場所：筑波大学、講師：滝沢直宏、参加者約10人。

13. ワークショップ「エディターを使ったコーパス検索」

開催日時：2011年3月7日～8日、開催場所：筑波大学、講師：滝沢直宏、参加者約10人。

14. 第1回複合辞研究会

開催日時2007年12月16日(日)、開催場所：筑波大学
発表5件、参加人数：約30名。

15. 第2回複合辞研究会

開催日時：2009年2月22日（日）開催場所：筑波大学
発表7件、参加人数：約30名。

16. 第3回複合辞研究会

開催日時：2010年2月13日（土）開催場所：筑波大学
発表7件、参加人数：約30名。

17. 第4回複合辞研究会

開催日時：2011年1月29日（土）開催場所：筑波大学
発表4件、参加人数：約30名。

辞書編集班：

18. 辞書編集班拡大班会議（テーマ：コロケーション研究会）

開催日時：2007年11月10～11日、開催場所ホテルスワ（つくば市）

発表15件、参加人数：25~30名

19.辞書編集班拡大班会議（テーマ：コーパスと辞書編集）

開催日時：2008年11月23~24日、開催場所ホテルスワ（つくば市）

発表15件、参加人数：25~30名

20.辞書編集班拡大班会議（テーマ：コーパスと辞書）

開催日時：2009年11月14~15日、開催場所ホテルスワ（つくば市）

発表15件、参加人数：25~30名

21.辞書編集班拡大班会議（テーマ：コーパスと辞書）

開催日時：2010年11月13~14日、開催場所ホテルスワ（つくば市）

発表15件、参加人数：25~30名

C) 招待講演・基調講演の類

領域代表者による国際学会・ワークショップでの本領域の活動についての招待講演・基調講演を示す。国内のワークショップについては9.1A)参照。

1. Kikuo Maekawa. “KOTONOHA and BCCWJ: Development of a Balanced Corpus of Contemporary Written Japanese.” *Corpora and Language Research: Proceedings of the First International Conference on Korean Language, Literature, and Culture*. Seoul, Yonsei University, pp.158-177, 2007:02.
2. Kikuo Maekawa. “Design of a Balanced Corpus of Contemporary Written Japanese.” *Proceedings of Symposium on Large-Scale Knowledge Resources (LKR2007)*, pp.55-58, 2007:03.
3. Kikuo Maekawa. “Balanced Corpus of Contemporary Written Japanese.” *Proceedings of the 6th Workshop on Asian Language Resources (ALR)*, pp.101-102, 2008:01
4. Kikuo Maekawa. “Compilation of the Balanced Corpus of Contemporary Written Japanese in the KOTONOHA Initiative.” *Proceedings of the Second International Symposium on Universal Communication*, pp.169-172, 2008:12
5. Kikuo Maekawa. “KOTONOHA: A Corpus Compilation Initiative at the National Institute for Japanese Language.” *Keynote speech at the 22nd International Conference on the Computer Processing of Oriental Languages*, Hong Kong Polytechnic University, Hong Kong, 2009:03
6. Kikuo Maekawa. “Development of Japanese Corpora at the National Institute for Japanese Language and Linguistics” *Keynote speech at Asialex 2011*: 2011:08(予定).
7. Kikuo Maekawa. “Linguistics-Oriented Language Resource Development at the National Institute for Japanese Language and Linguistics.” *Keynote speech at Oriental COCOSDA 2011*, 2011:10 (予定).

6. 4 「国民との科学・技術対話」

前項で述べたように、平成23年3月14~16日の三日間にわたって都心で公開ワークショップ（研究成果報告会）を開催する準備を進め、約300名の事前参加申込を得ていたが、3月11日の東日本大震災により会場が使用不能となり開催を中止した。会場費等の払戻金が生じたので平成23年度に繰り越した。その資金を用いて同年8月2,3日に「『現代日本語書き言葉均衡コーパス』完成記念講演会」を都心で開催する準備を進めている。

7. 研究組織と各研究項目の連携状況

7. 1 研究組織

総括班

前川喜久雄	領域代表者	国立国語研究所言語資源研究系	教授
山崎 誠	データ班長		
松本裕治	ツール班長		
傳 康晴	電子化辞書班長		
田野村忠温	日本語学班長		
砂川有里子	日本語教育班長		
田中牧郎	言語政策班長		
荻野綱男	辞書編集班長		
奥村 学	言語処理班長		
天野成昭	愛知淑徳大学	教授 (外部評価委員)	
後藤 斉	東北大学	教授 (外部評価委員)	
佐藤 宏	小学館	取締役 (外部評価委員)	
ソ・サンギュ	延世大学	教授 (外部評価委員)	
田中穂積	中京大学	教授 (外部評価委員)	平成21年度まで
古井貞熙	東京工業大学	教授 (外部評価委員)	
宮島達夫	国立国語研究所	名誉所員 (外部評価委員)	
木戸冬子	東京大学大学院情報理工学系研究科	(広報担当)	

計画班 (※印は班長、以下同様)

データ班：代表性を有する現代日本語書籍コーパスの構築

山崎 誠※	国立国語研究所言語資源研究系	准教授
丸山岳彦	国立国語研究所言語資源研究系	助教
柏野和佳子	国立国語研究所言語資源研究系	准教授
佐野大樹	国立国語研究所コーパス開発センター	特別奨励研究員
山口昌也	国立国語研究所言語資源研究系	助教
間淵洋子	国立国語研究所コーパス開発センター	特別奨励研究員
高田智和	国立国語研究所理論・構造研究系	准教授
小椋秀樹	国立国語研究所言語資源研究系	准教授
小沼 悦	国立国語研究所コーパス開発センター	職員

ツール班：書き言葉コーパスの自動アノテーションの研究

松本裕治※	奈良先端科学技術大学院大学情報科学研究科	教授
徳永健伸	東京工業大学大学院情報理工学研究科	教授
乾健太郎	東北大学大学院情報科学研究科	教授
橋田浩一	産業技術総合研究所社会知能技術研究ラボ	ラボ長
浅原正幸	奈良先端科学技術大学院大学情報科学研究科	助教
橋本泰一	東京工業大学総合プロジェクト支援センター	助教
飯田 龍	東京工業大学大学院情報理工学研究科	助教
小町 守	奈良先端科学技術大学院大学情報科学研究科	助教

電子化辞書班：多様な目的に適した形態素解析システム用電子化辞書の開発

傳 康晴※	千葉大学文学部 教授
峯松信明	東京大学大学院新領域創成科学研究科 准教授
小木曾智信	国立国語研究所言語資源研究系 准教授
小磯花絵	国立国語研究所理論・構造研究系 准教授
山田 篤	京都高度技術研究所研究開発部 主席研究員
内元清貴	情報通信研究機構総合企画部 主任研究員
小澤俊介	名古屋大学大学院情報科学研究科

日本語学班：コーパスを用いた日本語研究の精密化と新しい研究領域・手法の開発

田野村忠温※	大阪大学大学院文学研究科 教授
服部 匡	同志社女子大学表象文化学部 教授
杉本 武	筑波大学大学院人文社会科学研究科 教授
石井正彦	大阪大学大学院文学研究科 教授

日本語教育班：代表性を有する書き言葉コーパスを活用した日本語教育研究

砂川有里子※	筑波大学大学院人文社会科学研究科 教授
井上 優	麗澤大学大学院言語教育研究科 教授
小林ミナ	早稲田大学大学院日本語教育研究科 教授
滝沢直宏	名古屋大学大学院国際開発研究科 教授
投野由紀夫	東京外国語大学大学院地域文化研究科 准教授
山内博之	実践女子大学文学部 准教授
千葉庄寿	麗澤大学外国語学部 准教授
橋本直幸	首都大学東京オープンユニバーシティ 助教
曹 大峰	北京日本学研究中心 教授
小西 円	早稲田大学大学院日本語教育研究科
清水由貴子	早稲田大学日本語教育研究センター
本田ゆかり	東京外国語大学大学院地域文化研究科
奥川育子	筑波大学人文社会科学研究科

言語政策班：言語政策に役立つ、コーパスを用いた語彙表・漢字表等の作成と活用

田中牧郎※	国立国語研究所言語資源研究系 准教授
相澤正夫	国立国語研究所時空間変異研究系 教授
齋藤達也	専修大学文学部 准教授
棚橋尚子	奈良教育大学教育学部 教授
近藤明日子	国立国語研究所コーパス開発センター 特別奨励研究員
河内昭浩	群馬県立館林高等学校 教諭
鈴木一史	東京大学教育学部附属中等学校 教諭

辞書編集班：コーパスを利用した国語辞典編集法の研究

荻野綱男※	日本大学文理学部 教授
近藤泰弘	青山学院大学文学部 教授

矢澤真人	筑波大学大学院人文社会科学研究所 教授
丸山直子	東京女子大学現代教養学部 教授
荻野孝野	日本システムアプリケーション
星野和子	
橋本 修	筑波大学
魚 秀禎	啓明大学校
黒橋禎夫	京都大学大学院情報学研究科 教授

言語処理班：代表性のあるコーパスを利用した日本語意味解析

奥村 学※	東京工業大学精密工学研究所 教授
白井清昭	北陸先端科学技術大学院大学情報科学研究科 准教授
新納浩幸	茨城大学工学部 准教授
竹内孔一	岡山大学大学院自然科学研究科 准教授
佐々木稔	茨城大学工学部 講師
中村 誠	北陸先端科学技術大学院大学情報科学研究科 助教
高村大也	東京工業大学精密工学研究所 准教授
古宮嘉那子	東京農工大学工学研究院

公募班

機能表現班：（第2期）

宇津呂武仁※	筑波大学大学院システム情報工学研究科 准教授
--------	------------------------

作文支援システム班：（第1期、第2期）

仁科喜久子※	東京工業大学留学生センター 教授
--------	------------------

意見情報班：（第2期）

関 洋平※	筑波大学大学院 図書館情報メディア研究科 准教授
-------	--------------------------

日本語フレームネット班：（第1期、第2期）

小原京子※	慶應義塾大学理工学部 教授
-------	---------------

クラスタリング班：（第1期）

新納浩幸※	茨城大学工学部情報工学科 准教授
-------	------------------

リーダービリティ班：（第1期）

柴崎秀子※	長岡技術科学大学工学部教育開発系 教授
-------	---------------------

7. 2 項目間の連携状況

A)項目間の連携

本領域における項目設定は単純であり、A01「コーパスの構築」とB01「コーパスの評価」の2項目だけである。そのため項目間の連携関係も非常にシンプルであり、A01で構築したデータを利用してB01が各種の研究を実施し、そのなかで発見した問題点があれば、それをA01にフィードバックするという構図が想定されていた。そのため、領域の運用にあたっては計画期間の前半ではA01活動の重点をおいてデータの早期提供につとめ、計画期間後半ではB01に重点を移すことに努めた。

本領域の申請書(p.8)には「構築しうるコーパスデータ量の推移の予測」についての記述があり、図4が示されている。この図では「ベース」となる『日本語話し言葉コーパス』750万語に本領域で作成する書籍データ（図の「特定」）および国立国語研究所のKOTONOHA計画で作成する雑誌・新聞・ウェブ等のデータ（「国語研」）を合計して、平成19年度末に合計で4000万語弱相当、21

年度末に1億語超相当のデータを利用可能とする予定をたてていた。実際の運用実績はこの予定から極端にはずれずるものではなかったが、平成21年から22年にかけては一時的に予定を下回るペースとなった。その原因は、著作権処理が容易なサンプルがほぼ出尽くしてしまい、著作権処理に多大な時間がかかりだしたことで、3節に述べた国立国語研究所の移管問題が平成19年末に生じ、21年9月まで移管を終えるまで、業務に悪影響を及ぼしたためである。

この問題を解消するために21年度以降は著作権処理が未了のデータも領域内の研究者に限っては利用可能とし、研究に利用可能なデータ量の確保に努めた（ただしデータベース上ではそれが著作権処理未了であることを識別可能として、例文などには利用しないよう注意を促した）

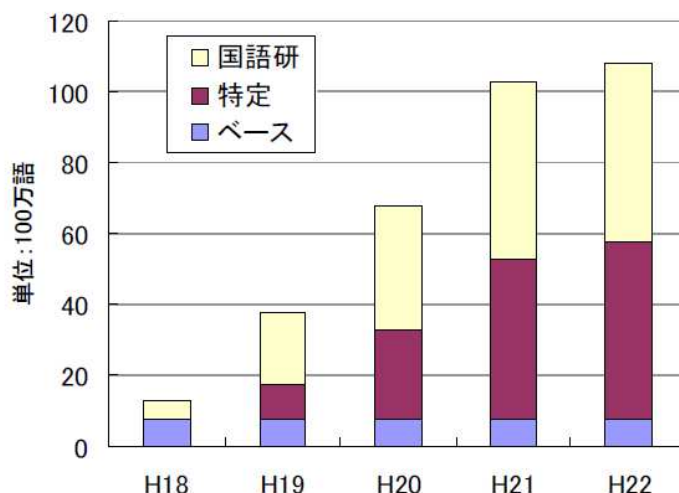


図4: コーパスデータ量の推移の予測
(本領域申請書より引用)

もちろん、いくらデータが提供されても、それを研究にうまく活用するためにはデータの性質を正しく理解する必要があり、また種々のコンピュータソフトウェアの利用についてのスキルも必要になる。この方面での問題を解消するために、データ班による『現代日本語書き言葉均衡コーパス』仕様説明会（2回）、ツール班による講習会（8回）を開催した（詳細は6.3節参照）。最大で50名ほどの参加者があり、項目間の連携に有益であった。

B) 計画班と公募班の連携

第1期公募班のうちクラスタリング班は研究手法および目的の両面において計画班のひとつである言語処理班と重なるところが多かったので、第4年度からは言語処理班の一部に吸収して活動を継続させた。これによって第2期公募研究の採用枠を実質上ひとつ増やすことができた。

日本語フレームネット班の活動も言語処理班の研究の一部と関係が深かったので、両者のデータをネット上で相互参照可能とし、研究の円滑な推進を図った。

作文支援システム班が構築した「なつめ」のために必要な論文データの著作権処理の一部については総括班（領域代表者）が協力した。

ツール班、言語処理班、日本語フレームネット班、意見情報班はそれぞれの研究目的を達成するために『現代日本語書き言葉均衡コーパス』の一部に独自のアノテーション（研究用情報付与）を実施したが、その対象とするサンプルができるかぎり重複するように総括班会議で調整した。その結果、『現代日本語書き言葉均衡コーパス』の一部には、データ班で付与する形態論（形態素）情報に加えて、述語項構造、統語構造、意味役割、評価情報など、多彩な研究用情報を付与することができ、コーパスの利用価値を高めることができた。

意見情報班の活動にデータ班の研究者が協力して評価情報の言語学的な理論化を担当した。

8. 研究費の使用状況

本領域ではデータ班によるコーパス構築に予算を大きく配分しているが、そこではコーパスの構築作業の進展を予測して予算を傾斜配分することを試みた（各年度の配分は、5430万円、8620万円、5490万円、2930万円、1750万円）。傾斜配分には良い面と悪い面があった。良い面としては実質的な活動期間が半年程度にとどまる初年度の予算を抑えたことによって研究費の無駄な執行を抑制できたことがあげられる。

悪い面としては第4、第5年度の予算を少なく見積り過ぎたために予算面で窮屈な運営を強いられたことがあげられる。ただし幸運なことに本領域は中間評価において最高評価（A）を得たことにより、これらの年度に予算の追加配分を受けることできた。追加予算はデータ班と言語政策班に配分したが、データ班においては上述の傾斜配分の問題を解消するために大変有効であった。

次にデータ班における予算の効率的使用について述べる。データ班ではコーパス構築のために大量のテキスト入力作業と文書構造などのアノテーション作業を実施した。このうちテキスト入力は民間業者に外注して実施したが、アノテーション作業は国立国語研究所において実施することとし、そのために任期付研究員と研究補佐員を合計10名程度雇用した。研究員、研究補佐員の人件費は民間業者に比べて割高であるから、単純に考えれば非効率的に見える可能性がある。

しかし前例のない高精度、高付加価値のコーパスを構築しようとするデータ班の活動には日本語の言語資源学の基礎研究としての色彩が強く、データをつくりながら同時にコーパスの仕様の細部を決定していく必要があった。このような作業を実施するためには仕様を考案する研究者とその指示に基づいて作業を進める研究補佐員とが同じ職場で随時意見を交換しながら作業を進める体制が必要不可欠である。仮にアノテーション作業を外注した場合、マニュアルの変更が契約の変更に直結することから、必要な改定を行うことができず、結果として不満足なコーパスを作成してしまう可能性が高い。

ちなみに国立国語研究所の移管問題（3.2節参照）の発端となった行政減量・効率化有識者会議の有識者達は『現代日本語書き言葉均衡コーパス』の構築は民間で可能と断言しているが、これはコーパスと一般の国語辞書を同一視したとんでもない誤認識である（詳しくは領域代表者のホームページ参照 http://www2.ninjal.ac.jp/kikuo/ikan_record.html）。

なおデータ班が雇用した研究補佐員には大学院教育をうけた者も多くふくまれていたが、彼らは研究期間の途中からはほぼ自律的にマニュアルの作成や改定を実施できるように成長し、公開研究成果発表会や一般の学会で研究発表を行う者も少なくなかった。そのなかから国立大学の常勤講師の職を得るものがでるなど、結果的に若手研究者の育成策としても有効であった（10節参照）。

領域全体での予算の効率的利用の例としては、領域で共有サーバを保有して、各班が様々な目的に利用したことが挙げられる。電子化辞書班のUniDicの公開サイト、日本語学班が領域内に限定して公開したウェブ文書のコーパスなどは、いずれも共用サーバを利用して実装されている。

最後に各班とも必要に応じて科研費の繰越制度を利用するなどして研究費の効率的利用を心がけたことは言うまでもない。

9. 当該学問分野及び関連学問分野への貢献度

9. 1 シンポジウム・特集・報道

既に2節に述べたように、本領域の活動に直接影響されたみなせるシンポジウムないし学会誌特集号が関連学会等により多数企画された。以下に一覧表を掲載する。いずれの機会においても本領域代表者が発表ないし寄稿しているが、とりわけ『人工知能学会』の特集「日本語コーパス」は本領域の活動の紹介そのものを目的とした特集であり、計画班の班長が各班の活動を紹介した論文（8篇65頁）から構成されている。

A) 学会等のシンポジウム

1. 日本言語学会第134回大会シンポジウム「大規模コーパス研究の方法—言語研究の新しいス

- タンダードの構築に向けて―」2007年6月 麗澤大学.
2. 英語コーパス学会第30回記念大会シンポジウム「他言語コーパス研究の現在：英語研究への示唆」2007年10月 立教大学.
 3. 情報処理学会Web DB Forum「企業の巨大データ徹底解剖：新たな研究の可能性と産学連携」2008年 学習院大学.
 4. 漢字文献情報処理研究会公開シンポジウム「著作権をめぐる最近の動向」,2009年7月 慶応大学日吉キャンパス.
 5. 言語資源シンポジウム「言語・音声データの学術利用に向けて」2009年 国立情報学研究所.
 6. 日本心理学会ワークショップ「人の語彙システム研究の多面的アプローチ(2) ―意味のコーディングを考える―」2011年9月予定, 日本大学.

B) 雑誌の特集

1. 専門誌『日本語科学』22号 特集「コーパス日本語学の射程」2007年.
2. 学会誌『日本語の研究』4巻1号 特集「資料研究の現在」2008年.
3. 学会誌『英語コーパス研究』特集「他言語コーパス研究の現在:英語研究への示唆」2008年.
4. 一般誌『国文学解釈と観賞』特集「日本語研究とコーパス」2008年.
5. 専門誌『漢字文献情報処理研究』特集「著作権をめぐる最近の動向 “Googleブック検索”著作権法改正案」2009年.
6. 学会誌『人工知能学会誌』,特集「日本語コーパス」2009年.
7. 学会誌『言語研究』138号 特集「コーパスを活用した言語研究」2010年.

C) マスコミ・ネットでの報道

1. @IT,「コトバのインフラ整備、著作権処理で法外なコスト 国立国語研究所が大規模コーパスを試験公開」2007年5月28日.
2. Internet Watch,「国立国語研究所、1,000万語分の日本語コーパスを試験公開 第1弾は「Yahoo!知恵袋」と政府刊行白書のデータ、最終的には1億語が目標」2007年5月28日.
3. 日経パソコンオンライン,「「風景」と「光景」の違いは? 国立国語研が大量の日本語データ公開」2007年5月28日.
4. MYCOMジャーナル,「現代日本語の書きことばの特徴が分かる 国語研がサイトを試験公開」2007年5月29日.
5. 日刊工業新聞,「日本語書き言葉のDB化 ブログの表記収集へ 国立国語研究所」2007年5月29日朝刊11面.
6. NHKラジオ第一,「NHKジャーナル」2007年6月4日.
7. 朝日新聞,「一千万語の書き言葉、ネット公開」2007年6月15日朝刊23(科学)面.
8. YOMIURI PC,「誰が日本語を作ったか(山野邊一也)」2007年8月号.
9. 毎日新聞ユニバーサロン,「音声コーパス、著作権法が壁—NIIがSRCシンポジウム」2007年11月15日.
10. 東京MXテレビ,「ガリレオチャンネル」2008年2月10日「言葉の壁を越えて:音声翻訳技術が世界を変える」
11. Internet Watch,「現代日本語の研究などに「Yahoo!ブログ」の記事を収集・提供へ」2008年4月23日.
12. IT Pro「NIIと国語研、Yahoo!ブログのデータを研究用途で無償提供」2008年4月23日.
13. IT Media News,「「w」(ワラ)の意味も解析!?!—NIIと国語研、Yahoo!ブログの記事を学術利用」2008年4月23日.
14. 日経パソコン,『Close Up 「人知の限界」を補う一億語の日本語データベース』2009年11月23日.
15. Research Map『つながるコンテンツ 研究の壁を越えたときX「これは壁ではない」』2011年3月1日.

9. 2 『現代日本語書き言葉均衡コーパス』モニター版利用者

『現代日本語書き言葉均衡コーパス』モニター版申込者（2009年版で834名）の研究上の背景を分類すると、狭義の言語研究が65%、言語教育（日本語教育と国語教育）が25%、残り10%が情報系である。情報処理系からの申請数は多くはないが、Apple社のiPadやiPhoneに搭載されている仮名漢字変換辞書がUniDicをもとに構築されているように、実社会に広く流通しているアプリケーションも既に出現している。

10. 研究計画に参加した若手研究者の成長の状況

本領域に参加した若手研究者には以下のように著しい成長が認められた（一部に本年9月以降の採用・昇任予定を含む）。今後のコーパス日本語学を担う人材の養成に大きく貢献できた。

- A) 本領域参加時、助教であった者5名のうち、3名が同一機関の准教授に昇任し（東京工業大学1名、国立国語研究所2名）、1名が他機関（国立国語研究所）の任期付准教授に採用される予定である。
- B) 本領域でポスドク（および国立国語研究所の奨励研究員）として雇用した者6名のうち、任期付常勤講師に採用された者が1名（愛知学院大学）、任期付助教となった者が1名（東京工業大学）、独立行政法人任期付研究員となった者が1名ある（NICT）。他に国立国語研究所の研究補佐員から常勤講師に採用された者が1名ある（群馬大学）。
- C) 本領域に参加した多数の大学院博士課程学生のうち、学位取得後に大学准教授に採用された者が1名（スロベニア国リュブリャナ大学）、助教に採用された者が2名（奈良先端大、東京農工大）、ポスドクに採用された者が2名ある（奈良先端大、筑波大学）。他に本領域に関係した研究テーマで学位を取得した者が4名ある（早稲田大学、奈良先端大、日本大学2名）。
- D) 本領域に協力者として参加した高校教師1名が大学専任講師に採用された（安田女子大学）。
- E) 博士課程学生から助教に採用されたものが1名（奈良先端大）、ポスドクに採用されたものが2名あった（奈良先端大、筑波大学）。

11. 総括班評価者による評価の状況

本領域では7名の専門家、有識者に外部評価委員を委嘱して、各年度末の公開ワークショップ（研究成果発表会）の折に評価委員会を開催した（評価委員の氏名、所属は7.1節の総括班メンバー表参照）。多くの委員は例年9月に開催した領域全体会議（非公開）にも参加した。

外部評価委員会では毎年活発な意見交換が行われた。主な指摘を以下に列記する。

- 1) 領域活動期間の早い時期からコーパスの応用を考えておくこと
- 2) コーパス応用の成功例を示すこと
- 3) データをタイムリーに公開して政府の言語政策にも影響を及ぼすこと
- 4) 領域外の研究者をまきこんだ活動とすること
- 5) 領域外の人間に成果がみえるようにすること

これらの意見のすべてに十分に対応できたわけではないが、『現代日本語書き言葉均衡コーパス』のモニター公開の実施、公開ワークショップへの公募サテライトセッションの導入によって、3)、4)、5)には対応できたと考えている。3)については、常用漢字表の改定を検討していた文化審議会国語分科会漢字小委員会に『現代日本語書き言葉均衡コーパス』の一部を参考データとして提供した。また3節で言及した文化審議会著作権分科会法制小委員会での参考意見聴取も3)にかかわる活動であった。1)と2)については十分に対応できていないが、言語政策班が作成した教育用語彙表は国語教育界に大きな反響を呼び起こしつつある。