様式 C-19

科学研究費補助金研究成果報告書

平成 22 年　3 月 31 日現在

研究種目：基盤研究（B）
研究期間：2007～2009 年
課題番号：19320081
研究課題名(和文)「コーパスをもとにした語学教育 e-Learning コンテンツの開発環境の研究」
研究課題名（英文）“Studies of development environment of language e-learning contents based on corpora”
研究代表者
佐野　洋
東京外国語大学・大学院総合国際学研究院・教授
研究者番号：30282776

研究成果の概要（和文）：本研究では、コーパスを利用した語学教育開発にあたり、質的なコーパスと量的なコーパスの双方の利用環境を整えた。(1) 日本語 e-Learning コンテンツの開発環境の基礎資料として、日本語教科書コーパスを開発した。(2) 英語 e-Learning コンテンツの分野では、句動詞用例のコーパス化と用例サイトの作成を行った。(3) 語学教育にウエブ上の実用例を活用するために、汎用コーパスである Google 5gram 英語データ（1 兆語）から用例を自動抽出する検索システムを構築した。

研究成果の概要（英文）：For the purpose of language education, this study has realized the environment for usage of both quality and quantity corpora. (1) We have developed a corpus of Japanese language for analytical use based on Japanese school textbook data. (2) We have created a corpus of English phrasal verbs and its website for student study. (3) We have made an automatic retrieval system optimized for Google 5gram corpus, which contains one trillion English words in it.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2007 年度	3,900,000	1,170,000	5,070,000
2008 年度	4,400,000	1,320,000	5,720,000
2009 年度	3,500,000	1,050,000	4,550,000
年度			
年度			
総　計	11,800,000	3,540,000	15,340,000

研究分野：人文学
科研費の分科・細目：言語学・外国語教育
キーワード：教材・教具論、コーパス

１．研究開始当初の背景

これまでにコーパスからの用例収集(及びウェブサイトを通じた用例の提供枠組み)と、語学教育のための e-Learning システムの研究開発を行ってきた。
(1) コーパス検索手段の研究(検索手段の改良と検索速度の高速化)
自然言語処理技術を使い、コーパスから英文を抽出することで多くの用例を得ることができる。佐野(研究代表者)は、(株)小学館と共同研究を続け、英語コーパス(BNC：British National Corpus(http://www.natcorp.ox.ac.uk/))から、英語用例を収集し、1320 項目からなる文法

項目別用例集(小学館コーパスネットワークから公開中。http://www.corpora.jp/)と、4319の見出し語からなる句動詞用例集を試作し、語学教育教材を作成してきた。コーパス検索言語はCQL(Corpus Query Language (株)小学館が開発)といい、単語を、抽象化された単位(表層、見出し、品詞標識)で表現し、単語連鎖パターンを使って用例検索を行う。
not only ... but also のように文型パターンが形態レベルの単位で表現できる場合は直接抽出できるが、他動詞構文(主語、目的語など構文レベルの抽象化された単位)のような文型パターンは、積(and),和(or),差(diff)の集合演算を組み合わせることで目的の用例を抽出する。ひとつの語彙・文型について、少ない場合でも 2〜3、多い場合では 20 以上のCQL 式と複数の集合演算式を記述する。
語彙・文型の運用知識を充実させるには、これまでの100倍以上の検索式を記述する必要があり(後述)、言語知識の表現レベルと、CQLによる検索式記述の表現レベルの乖離が大きく、式記述や管理が煩雑になるなど構文レベルの抽象化単位で検索式を記述する必要性が生じてきた。
また、1320 項目(文法項目別用例)の場合、すべての項目の検索に約 5 日間を要し、句動詞用例(4319 見出し)では約 20 日を要する。従って、これまでの 100 倍以上の規模の検索式で用例検索を、実時間で実施する仕組みが必要である。本研究では、DI(Dependency Injection)などの検索式間の依存性を管理する手法による検索式のオプティマイズと、ディスクを使わずメモリー上でデータを処理する検索エンジンの改良の、2 方式の組み合わせによって高速化を実現する。

２．研究の目的

語彙(語義)や文型についての概念形成(いわゆる語感獲得)には、練習用例数は多いほどよいといわれている。一般に、語学の文法学習書や辞書では、学習項目毎に、あるいは語彙(語義)毎に、関係する用例が、紙媒体という記録制約もあって１例しか掲載されていないことが多い。電子教材や電子辞書も、データ作成時の参照先が文法学習書や辞書なので、用例が少ない事実は変わらない。
一つに絞り込まれた典型用例が示す概念を知識とするだけでは、実際的な言語の運用はできない。語彙や文型には、さまざまな側面があり、基底の意味(core meaning)だけでなく、意味・用法拡張の方向も複雑で、他の単語との連鎖・共起も多様である。この言語知識(運用知識)を知っていれば認知力が働いて、単語や文型の持つ概念の多様さ、深さ、意味推移まで含めたことばの効果的な運用が可能になる。
語彙や文型についての概念形成には、練習用例数が多いほどよいのだが、一方、外国語として言語を学ぶ環境では、学習時間の制約があるので、言語運用知を効率的に提示する仕組みによって、学習効率を向上しなければならない。
本研究では，各語彙・文型の運用知識をコーパスから収集して組織化し，それら用例と概念視野を拡大する手段を提供することによって、語学教育の高度化と効率化を実現する。語学教育教材(とくに e-Learning 用コンテンツ)の作成に言語知識を活用するための理論及びシステムを開発することを目的とする。
この研究課題は以下の項目である。
(1) コーパス検索手段の研究(検索手段の改良と検索速度の高速化)
(2) コーパスからの用例収集と組織化の研究
(3) 運用知識の視覚化を含む、独自の e-Learning コンテンツ開発環境の研究

３．研究の方法

研究組織は，代表者の佐野が全体を統括し，データベースシステム班，用例収集分析班，コンテンツ開発環境・教材班の三つの班で構成する。

【各研究班の役割】

データベースシステム班は、(1)新しい検索式の言語設計を行う(言語知識の表現レベルと、現 CQL による検索式記述の表現レベルの乖離が大きく、式記述や管理が煩雑になっており、その問題を解消することが目的)、(2)新しい検索言語を CQL へ翻訳するコンパイラ、及び CQL レベルでのオプティマイザー(検索式間の依存関係を利用する)を開発する、(3)インメモリー方式のコーパスデータ検索システムを構築する。(4)既存の日本語コーパスと教材用例を基に、用例収集分析版と協力し、CQL を使って検索ができる日本語コーパスのデータベース構築を行う。
用例収集分析班は，(1)既存の検索式と使って用例検索を行う。また、データベースシステム班が設計する新しい検索言語を用いて検索式を作成し、用例抽出を行う、(2)抽出した用例について、中心語と各パターンで収集された結果の関係を、頻度情報や共起強度などの基本統計量を求め分類する、(3)言語の使用の偏りの分析から言語運用知識を組織化する。
コンテンツ開発環境・教材班は，(1)言語使用の偏りの分析結果と、検索属性(語形、文型、モーダル、主節形、従属形、学習レベル、語彙水準など)を使って、用例への自動タグ付けを行う、(2)言語運用知識を視覚的に提示する手法を検討し、学習項目に対応する１つの制約(目的別解釈)と、学習対象の中心語の他の語との共起制約など、言語が持つ多重の制約(体系的解釈)を多視点的に提示するインタフェースを実現する、(3)e-Learning 補助教材としての評価や、用例学習サイトと

しての評価を通じて、手法の妥当性の検証を行う。

４．研究成果

４．１　平成 19 年度

本年度(2007 年度)、日本語コーパスの整理を実施した。教科書コーパスの日本語用例文を収集、整理した。日本で販売されている 1400 冊の教科書の内、指導書等の電子データのある教科書を対象にコーパス収集を実施した。また、日本語述部の定形/非定形表現の基礎的な種類の特定を行い、自動分析するために形態素解析結果から、基礎的な定形/非定形表現を生成するプログラム等を開発した。

英語用例のコーパス化と用例サイトの作成を行った。冠詞や仮定法などと共に句動詞(動詞＋前置詞/副詞)は、日本人学習者にとって学習(理解と習得)が難しい文法項目であることが指摘されている。困難さの 学習パラメータとして、句動詞を構成する不変化詞(前置詞/副詞)に、動的な 意味があること、及び視覚的なイメージを想起させる機能があることを前提に、句動詞用例サイトを試作した。グループ化された前置詞/副詞が持つ機能的なカテゴリに着目し、そのカテゴリの認知的な解釈イメージを視覚的に示すことで、句動詞の意味理解を促進し、学習効果を高める英語教材提供のためのウェブサイトである。機能的なカテゴリは、動作を示す時空間イメージや、因果関係、状態変化などのイメージからなる。動詞の語彙的な意味、前置詞/副詞の意味に加えて、グループ化された前置詞/副詞が持つ機能的なカテゴリの 3 つの視点を意識させる学習用例は、学習効果の改善をもたらすと考えられる。2008 年度は評価を実施する予定である。

４．２　平成 20 年度

本年度(2008 年度)、昨年度に整理を行った日本語コーパスの分析を実施した。コーパス（日本で販売されている 1400 冊の教科書の内、指導書等の電子データのある教科書から収集）を

使って、日本語教科書語彙分布調査を行い、語(形態素、単語、文末語形)について、頻度情報を求め、累積比率を計算した。高頻度語の言語的な特徴について調べたほか、統計的な累積ポイントである、2σ、3σや 4σ点の語と頻度について調査した。

タグ付きコーパスを管理・検索するためのツールとして、小学館で開発された検索エンジン(JSC)と、NAIST（松本研究室）で開発された Chaki（茶器）の利用を試みた。Chaki に教科書コーパスを実装し、検索可能な状態とした。

なお､JSC は､UNI-Code 対応にしているので、日本語だけでなく、他の言語でもタグ付きコーパスの形態であれば、JSC に実装することで検索が可能になる。例えば、英語であれば、BNC タガー(CROWS)を使って英文を形態素解析し、タグ付きのデータにすることで、検索対象のデータの作成が可能であることを確認した。

本年度、未使用予算が発生した。研究分担者との作業分担打ち合わせでの齟齬もあって、お互いに発注したとの認識で進めていたが、未発注であったことが 3 月末に判明したためである。本研究に対する影響は少なく、次年度に購入することでカバーできる範囲のものである。

４．３　平成 21 年度

本年度(2009 年度)、日本語コーパスの増強を行った。すでに集積しているコーパス（日本で販売されている 1400 冊の教科書の内、指導書等の電子データのある教科書 170 冊から収集）のカタログを再調査し、教科バランスと絶対的に不足していた科目（数学、英語、音楽等）の指導書を購入し(52 冊を追加し)、プログラムを用いてテキスト化を行った。テスト問題やドリルなどについてもテキスト化を行った。同時に、教科書データの特徴比較のため家電製品のマニュアル文章のテキスト化も実施し、両コーパスの文末表現の比較を行った。また、日本語教科書データの検索システムの改良を実施した。

英語教材作成のためのツールとして、Google 5gram 英語データ（1 兆語）を使った検索システムを構築した。このシステムは、RDB(Relational Data Base)を用いて検索を行う。用例検索では制限された正規表現を扱うことができる。なお、Google 7gram 日本語データ（2500 億語）についてはデータを購入し、英語の検索システムに準じて検索システムとして実装することができることを確認した。当初は、検索式の工夫による用例検索の高速化を目指したが、検索方法の工夫によって一般的なデータベースシステムだけで高速検索が実現した。

また、英語への翻訳難易度の視点から、日本語従属表現の定性分析を実施した。従来の翻訳指導書だけでなく日本語教育でも、意味解釈の複雑さが指摘されている、日本語の連用形について、英語翻訳のための類型の調査と、類型を判断するための言語テストの方法を確立し、日本語と英語の表現特徴が著しく異なる用例について 1,000 文余りの翻訳文を作成した。

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１件）

藤村知子、芝野耕司、佐野洋、藤森弘子(2009)「学習者の気づきを促すeラーニングの活用」『日本語教育学会秋季大会予稿集』pp.279-280, 日本語教育学会

〔学会発表〕（計 10 件）
佐野洋、「語学教育に特化した LMS(JPLANG)とその教材作成機能」、情報処理研究者集会、文部科学省,pp250-251(2 頁),2007 年 11 月 11 日

竹内豪、竹内まち子、荒川裕子、佐野洋、「認知的な視覚情報を利用した句動詞用例提示の一方法」、情報処理研究者集会、文部科学省,pp253-255(2 頁),2007 年 11 月 11 日

竹内まち子、荒川裕子、竹内豪、佐野洋、「認知的視覚情報を取り入れた句動詞学習教材の考察」、情報処理研究者集会、文部科学省,pp256-258(2 頁),2007 年 11 月 11 日

于壮飛、佐野洋、「日本語教材作成のための用例抽出システムの開発 - 日本語コーパス(JSP)の構築」、情報処理研究者集会(文部科学省)、pp240-243(2 頁),2007 年 11 月 11 日

竹内豪，竹内まち子，佐野洋、「Ajax技術を用いた語学教育用例検索インタフェースの開発」、(社)情報処理学会、コンピュータと教育研究会(第 93 回、IPSJ-SIG-CE93)（8 頁）, 2008 年 2 月 17 日

SANO, Hiroshi, "JPLANGWiki - Material Development Aid for JPLANG", (Full Paper), 7page, SITE2008 Conference (Society for Information Technology and Teacher Education International Conference), Las Vegas, Nevada USA, March 3-7,2008.

Tomoko Fujimura,"中級から上級への橋渡しの聴解授業－e-learning教材JPLANG「ミニ講義」について－ (Classes for raising lecture comprehension from the intermediate to advanced level: Using E-learning JPLANG mini-lectures materials)", 2008 ATJ Seminar （Association of Teachers of Japanese),Apr 3, 2008,Hyatt Regency, Atlanta, Georgia

佐野洋、藤村知子、芝野耕司、「JPLANGWiki：JPLANG用教材作成ツール」、ICJLE2008 (International Conference on Japanese Language Education 2008 日本語教育国際研究大会)、釜山外国語大学、釜山、July 11-13, 2008 年

竹内まち子、佐野洋、芝野耕司、"Evaluating Waystage and Threshold Vocabularies by the BNC and the Google Web 1T" American Association for Applied Linguistics (AAAL) Denver Marriott Tech Center, Denver, Colorado, 2009 年 3 月 21 日

藤村知子、佐野洋、芝野耕司、藤森弘子、「学習者の気づきを促すeラーニングの活用」、日本語教育学会秋季大会、九州大学、pp279-280(2 頁)、2009 年 10 月 11 日

佐野洋、「日本語教科書コーパスの構築」、2009 日本語教育国際研究大会、ニューサウスウェールズ大学（シドニー）、2009 年 7 月 13 日

佐野洋、「ハイテクな日本語」、『月刊日本語』10 月号特集、pp22-25(4 頁)、株式会社アルク、2008 年

６．研究組織
(1)研究代表者
佐野　洋（東京外国語大学・大学院総合国際学研究科・教授）

研究者番号：30282776

(2)研究分担者
芝野　耕司（東京外国語大学・アジア・アフリカ言語文化研究所・教授）
研究者番号：50216024
在間　進（東京外国語大学・名誉教授）
研究者番号：30117709
馬場　彰（東京外国語大学・名誉教授）
研究者番号：90033446
藤村　知子（東京外国語大学・留学生日本語教育センター・准教授）
研究者番号：20229040