

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 30 日現在

機関番号：34302

研究種目：基盤研究(C) (一般)

研究期間：2012～2015

課題番号：24520668

研究課題名(和文) Webから獲得した言語知識をベースとするインタラクティブな外国語学習法の開発

研究課題名(英文) Development of the Interactive Method for Foreign Language Learning Based on Linguistic Knowledge Extracted from the Web Corpus

研究代表者

野澤 元 (Hajime, Nozawa)

京都外国語大学・外国語学部・講師

研究者番号：30450696

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：大規模な英字新聞コーパスを用いて、どのような名詞が様々な動詞の主語や目的語として頻りに使われるのかを調べ、コーパスに見られるそういった共起する動詞と名詞のパターンを、動詞の用法に熟達する上で英語の学習者が習得する必要がある文法知識とみなした。また、コーパスから得られたそういった用法の例文を手作業で編集することによって、より自然で百科事典的な知識とも一致する、学習者のための文法問題を作成した。さらに、それらの文法問題をインタラクティブに練習することのできるゲーム形式のEラーニングシステムを開発した。

研究成果の概要(英文)：We conducted a survey with a large scale English newspaper corpus to see what types of nouns are frequently used for various verbs as their subjects and objects, and defined the patterns of the co-occurring verbs and nouns extracted from the corpus as the grammatical knowledge which learners of English need to acquire to master the use of verbs. We created grammatical exercises for learners which are more natural and consistent with encyclopedic knowledge, by manually editing the examples of such usages taken from the corpus. We also developed a gamelike e-learning system on which learners can interactively do the exercises.

研究分野：認知言語学

キーワード：教育工学 言語学 情報工学

1. 研究開始当初の背景

(1) 携帯情報機器を用いた自立型 E ラーニングを可能にする環境: 携帯情報機器の爆発的普及に伴い、それらを用いた外国語学習の試みも広く行なわれている。情報機器は学習者の達成度に応じた教材を、それぞれのペースに合わせて管理しながら提示することができる。また、特に携帯可能な情報機器は場所と時間に制約されない手軽な学習を可能にする。このように、携帯情報機器を用いた自立型 E ラーニングを可能にする環境は整いつつある。これまでも、情報機器を利用した様々な外国語学習の教材が開発されてきた。しかし、それらは音声面に焦点を置いたものであるか、もしくは、紙ベースの問題を電子化することで効率化したものが大半であった。しかし、このような E ラーニングの手法は情報機器の利点を十分に生かしていない。

大規模実データへのアクセス: 情報機器はコーパスのような大量のデータの収納と検索が可能であるため、言語用法についての圧縮した実世界を提供することができる。

学習者とのインタラクション: 学習にインタラクション性を導入するという点で、情報機器は有効である。そもそも言語使用は他者とのインタラクションであり、インタラクション性の導入は、より動機付けの高い学習を可能にする。

(2) 百科事典的知識(世界についての一般的知識)に基づく外国語学習: 認知言語学では、文の意味は語の意味の単なるつなぎ合わせではなく、語や構文が喚起する百科事典的知識と見なされる。言語の意味と一般的知識を区別しないアプローチによる文の理解の研究は、近年広く支持されているが、これを応用した外国語教育は本格的には行なわれていない。他方で、近年ではコーパスを利用した辞書編集の手法が広く用いられている。しかし、コーパスを用いた手法は、一般の外国語学習教材の作成には使われていない。例えば、文法問題の例文は、教材作成者の直感から作られたものが多く、現実の用法という点では不自然なものが少なくない。よく批判される "This is a pen." などという文は、実際に用いられる機会がほとんどなく、用法的に不自然であるため、学習者にとっては質の悪い学習刺激である。言語使用には無意識な側面があり、教材作成者は必ずしも自然な用法を意識できるわけではない。コーパスはこの問題を解決する手助けとなる。特に、Web コーパスから構築された集約的なデータは、人手によっては得ることができない自然さの程度や使用頻度といった情報を提供する。コーパスに基づくデータは、言語使用者が持つ一般的な知識を反映した、自然でより動機付けの高い例文の作成を可能にすることが見込

まれる。

(3) インタラクションとしての外国語学習: 認知言語学では、言語習得を語や文法の暗記ではなく、それらの表現を用いたインタラクションの体得と見なす、機能主義的な研究が行なわれている。会話を通じた外国語学習は、その典型的な様式であるが、一般的に行なわれている紙ベースの練習問題を解く学習は、インタラクション性が低く、学習の動機付けが弱いという問題がある。携帯情報機器を用いた E ラーニングの手法を用いることにより、よりインタラクション性の高い学習環境を実現することができる。

(4) 大規模コーパスからの言語知識の構築: 申請者らは、それぞれの語がどのような文脈で生起するののかについての情報を、大規模コーパスを用いた格フレーム辞書として自動構築した。この技術は、現実世界での言語用法を集約的に観察記述するために有用な手段である。

2. 研究の目的

(1) 百科事典的知識に基づく外国語学習教材の開発

言語と知識を一体化した外国語学習: 例えば、"People X turkey on Christmas day." という文のスロット X に、どのような語が生起するかは、文法的な問題であると同時に百科事典的知識の問題である。また、文法的に生起可能な語は無数にあるが、実際の用法には百科事典的知識と一致する限られた語のみが見られる。言語習得は、このような一般的知識に強く動機付けられた学習が行なわれるが、これは外国語学習においても同様である。これまでの日本の外国語教育は、語彙や文法を覚えることに力点が置かれ、習熟は会話などの実践を通して達成されると考えがちであった。しかし、言語習得は百科事典的知識に動機付けられており、一般的知識は学習戦略の重要な要因であるにもかかわらず、これを取り入れる試みは少なかった。本研究課題では、大規模コーパスを用いて、百科事典的知識に根ざした効果的な外国語学習教材を開発する。

語彙学習と構文学習の不可分性: 単語帳と文法書の二分化に見られるように、一般的に語彙と文法は別々の学習項目であると考えられている。しかし、ある語は常に特定の構文に埋め込まれて用いられ、語彙と文法の学習は不可分である。本研究課題では、古典的と見られがちな穴埋め問題を学習課題の形式として利用するが、これは語彙と構文を関連づける有効な形式である。これまでの穴埋め形式の学習課題は、抽象的な文法性の判断や慣用句の知識のみに焦点を置いていた点が問題であり、本研究課題では、コーパスが

らの実例に基づくことで、語彙と構文の一体的な学習を可能にする学習課題を作成する。

現実世界を反映した外国語学習: 語の用法は可能性全体から見れば偏りがあるが、その多様性自体は決して小さくなく、教材作成においては、人手作業だけでは十分に現実世界を反映した例文を用意できない。教材作成者の直感だけでは多様性の一部しか書き出せず、被覆率が不十分になる。また、用法の典型性や頻度の情報も人手では得られない。本研究課題では、この問題を解決する。

(2) インタラクションとして言語を習得する教材の開発

インタラクションとしての学習課題: 多くの外国語学習の教材は、学習者に抽象的な文法から見た「正しい文」を産出するよう促す形式をとる。しかし、「正しい文」の産出自体は、通常コミュニケーションの目的ではなく、このような形式は学習者を強く動機付けない。本研究課題では、百科事典的知識についての質問という形式で、インタラクションに埋め込まれた学習課題を作成する。

携帯情報機器による実装: 本研究課題では、学習課題の構造を工夫するだけでなく、CATシステムを利用しながら、携帯情報機器によってリアルタイムのインタラクションを生み出すことで、より自然な言語習得の環境を作り出すことを目指す。

(3) 教育と言語学と言語情報処理の融合: 理論言語学を教育に生かす試みはこれまでもあったが、必ずしも十分な成果をあげていない。その理由の一つは、理論言語学から提案される手法に、言語使用の現実的裏付けがなかったからである。本研究課題では、言語情報処理技術を利用することで、言語使用の実態を反映した教材の作成を目指す。これら三つの領域の融合により、新たな外国語教育の可能性を切り開く。

3. 研究の方法

(1) 研究方法の全体像: 研究は「問題アイテム作成班」と「システム構築班」に分かれて行ない、双方が互いの成果を基盤としながら連携して進めた。研究期間は「問題アイテム作成期」、「アプリケーション開発期」、「成果公開・発展期」の三つに分けられた。問題アイテム作成期には、問題アイテム作成班による予備調査に基づき、システム構築班が Web コーパスから言語使用の調査、抽出を行い、問題アイテム作成班がその結果を用いて英語学習のコアアイテムを作成する。アプリケーション開発期には、システム構築班がアプリケーションの開発と改良を行い、問題アイテム作成班がその実証実験とフィードバックの収集・分析を行なう。成果公開・発展期

には、作成した英語の学習アプリケーションを無償で公開し、社会的還元を行なうほか、他の外国語学習への応用を検討する計画であった。

(2) 問題アイテム作成期: 英語学習のコアアイテム作成のための調査とアプリケーションの基本設計を行った。

コアアイテムのための予備調査 <問題アイテム作成班>: 当初は日本十進分類法等を参考にして知識領域を区分し、それぞれの領域ごとに予備調査として入手可能な例文を収集する予定であった。しかし、実際に作業を始めると、Google など民間の無料検索サービスでは項構造を反映した検索には限界があり、他方で、Sketch Engine(民間の有料検索サービス)や BNC、COCA といったコーパスではコーパスのサイズが小さすぎたため、「百科事典的知識に基づく外国語学習教材の開発」という目標にとって有用なデータが収集できなかった。そのため当初の計画を変更し、はじめに知識領域を区分するのではなく、語彙ごとに例文を収集し、可能であればのちにそれらを知識領域に基づいて区分することにした。つぎに、システム構築班が独自に構築した大規模 Web コーパス(10 億文)を基にした、述語、述語・項のペア・インデックスを用いて、予備調査として特定の名詞キーワードを項とする動詞のリストとその例文、また、特定の動詞キーワードの項となる名詞のリストとその例文を収集しようと試みた。しかし、この Web コーパスはサイズの点では十分に大きいものの、文法的に正確な文や断片的な文が多く含まれており、さらに依存構造解析の精度も不十分であったため、予備調査に必要な検索結果を得ることができなかった。そこで、使用するコーパスを文法構造がより正確だと思われる新聞コーパスに切り替え、the English Gigword corpus の一部(280 万文)を用いて作成された、述語、述語・項のペア・インデックスによる予備調査を行った。これにより検索結果は一定程度の改善が見られたが、それでもなお依存構造解析の精度が不十分であり、コーパスサイズを増やしたとしても必要な情報がえられない可能性が高いことがわかった。最終的には、the English Gigword corpus の一部を用い、特定の動詞の主語と目的語として共起する名詞のペアのインデックスを、共起頻度と相互情報量を基にして作成し、それを用いて英語の動詞の用法を学習するコアアイテムを作成することになった。

問題アイテム作成を支援するための Web コーパス作成と格フレーム辞書構築 <システム構築班>: 英語を対象とした Web コーパス(10 億文)を、Web からクロールした Web ページ集合から 40 単語以下の英文を抽出することで作成した。また、これに対して

MaltParser による依存構造解析を行い、述語・述語・項のペア・インデックスを作成した。さらに、より精度の高い結果を出すために、the English Gigword corpus の一部(280 万文)を用いて、述語・項のペア・インデックスを作成した。また、依存構造解析の精度を上げるために、動詞とその項となる名詞の共起性から算出した語の親和性を用いて、構文の解釈が曖昧となる特定のパターンの文の非曖昧化の手法の開発を試みた。

(3) アプリケーション開発期: 英語学習のコアアイテムを作成し、またコアアイテムを提示するアプリケーションの構築とその評価の実証実験を行なう。

コアアイテムの作成 <問題アイテム作成班>: the English Gigword corpus の一部(280 万文)を用いて、共起頻度と相互情報量を基に作成された述語・項のペア・インデックスと、その例文を手で編集することによって、英語の動詞の用法を学習するためのコアアイテムを作成した。当初の計画では、特定の動詞に対して共起頻度の高い主語と目的語のペアを含む例文を自動的にコーパスより選別し、また、依存構造解析に基づいて自動的に問題文を作成する予定であった。しかし、精度の問題以外にも、自動的に選別された例文は、専門的な知識が必要で学習に適した文意でない、また、問題文としては長すぎるといった問題があった。そのため、当初の計画を変更し、担当者と作業補助者が特定の動詞毎に分類された例文のリストを基にして、手作業で例文を編集してコアアイテムを作成した。その際、ペアに含まれる名詞については JACET8000(石川・他 2005)を利用し、またコーパス中の頻度を考慮しながら、三段階の難易度に区分した。その際、ペアに含まれる名詞については JACET8000(石川・他 2005)を利用し、またコーパス中の頻度を考慮しながら、三段階の難易度に区分した。

アプリケーションの開発 <システム構築班>: 問題アイテム作成班が作成したコアアイテムを用いて外国語学習を行うためのアプリケーションを製作した。当初は、iPhone アプリとして製作を試みたが、アプリケーションの保守などの利便性を考慮し、最終的には HTML5 をベースとして開発した。

4 . 研究成果

(1) 英語の大規模コーパス内の共起頻度と相互情報量に基づいて作成された述語・項のペア・インデックスを利用して、百科事典的知識と一貫性のある動詞の用例を抽出し、また、その例文を手作業で編集して、英語の動詞の用法を学習するためのより自然なコアアイテムを作成した。このような手法で、210 題の問題が作成された。

(2) HTML5 をベースにしたゲーム形式のエラーニングシステムを構築し、英語の動詞の用法を学習するための問題を載せた。

(3) 依存構造解析の精度を上げるために、動詞とその項となる名詞の共起性から算出した語の親和性を用いて、構文の解釈が曖昧となる特定のパターンの文の非曖昧化の手法の開発を試みた。

5 . 主な発表論文等

[雑誌論文](計 2 件)

野澤元・河原大輔、コーパスから算出した語の親和性によって構文パターンの曖昧性を解消する試み、言語処理学会第 20 回年次大会発表論文集、査読無、20 巻、2014、pp.189-192

Daisuke Kawahara, Daniel W. Peterson, Octavian Popescu and Martha Palmer、Inducing Example-based Semantic Frames from a Massive Amount of Verb Uses、Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL2014)、査読有、Vol. 1、2014、pp.58-67

[学会発表](計 4 件)

李在鎬・赤瀬川史、均衡コーパスから抽出した基本動詞 100 語の格フレーム、日本語教育国際大会、2012 年 08 月 17 日~2012 年 08 月 20 日、名古屋大学(愛知県・名古屋市)

野澤元・河原大輔、コーパスから算出した語の親和性によって構文パターンの曖昧性を解消する試み、言語処理学会第 20 回大会、2014 年 3 月 18 日~2014 年 3 月 20 日、北海道大学(北海道・札幌市)

Hajime Nozawa and Daisuke Kawahara、A Self-correcting Approach to Solve Syntactic Ambiguities based on Collocational Strength、6th International Conference on Corpus Linguistics (CILC2014)、2014 年 5 月 22 日~2014 年 5 月 24 日、Las Palmas de Gran Canaria (Spain)

野澤元、認知言語学と外国語教育のデザイン、日本英文学会関西支部第 9 回大会、2014 年 12 月 21 日、立命館大学(京都府・京都市)

[その他]

ホームページ等

http://www.jlptreading.com/html5_quiz_game/#/70344550

6 . 研究組織

(1) 研究代表者

野澤元 (NOZAWA, Hajime)

京都外国語大学・外国語学部・講師
研究者番号：30450696

(2)研究分担者

河原 大輔 (KAWAHARA, Daisuke)
京都大学・情報学研究科・准教授
研究者番号：10450694

李 在鎬 (LEE, Jae-ho)
筑波大学・人文社会科学研究科・准教授
研究者番号：20450695