

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 11 日現在

機関番号：34310

研究種目：若手研究(B)

研究期間：2013～2015

課題番号：25870898

研究課題名(和文) 認知言語学に基づいた英語語彙・構文ネットワーク構築のための学習法開発

研究課題名(英文) Development of resources for English learners with special reference to the cognitive linguistic concept of the network of vocabulary items and linguistic constructions

研究代表者

長谷部 陽一郎 (Hasebe, Yoichiro)

同志社大学・グローバル・コミュニケーション学部・准教授

研究者番号：90353135

交付決定額(研究期間全体)：(直接経費) 1,700,000円

研究成果の概要(和文)：本研究は英語学習者が既存知識を最大限に活用しながら、さらに高度なスキルを身につけるための教材・学習法を開発を行うものである。認知言語学における用法基盤(使用依拠)モデルに基づいた研究・開発の結果として次の成果を得た。1) コーパス・データをもとに学習者にとって最適な英語構文のリストと対応する例文を作成した。2) 構文パターンとその実例を検索できるシステムTED Corpus Search Engineを構築した。3) 語彙・構文の知識を実践的な能力に結びつけられるよう、音声認識を利用したシステムE-Speakを開発した。これらの成果物は、単なる語彙学習にとどまらない総合的な英語学習の実践を促進する。

研究成果の概要(英文)：This project aimed to develop a set of resources and methodologies that enable learners of English to improve their skills based on what they have already built up in their preceding learning activities. The research and development in accordance with the theories of cognitive linguistics and the usage based model of language successfully produced the following: 1) a list of linguistic constructions in English with example sentences that optimize the learners' preexisting linguistic knowledge; 2) an web-based system called TED Corpus Search Engine (TCSE) that allows the learner to read and hear the words and phrases in the real context; 3) another web-based system utilizing a speech recognition engine which evaluates spoken English text input by the learner. These resources will promote a new way of learning English which focuses not only on vocabulary by itself but also on a complex network of linguistic constructions manifested in various forms in different contexts.

研究分野：認知言語学

キーワード：認知言語学 英語教育 語彙学習 構文学習 意味ネットワーク Eラーニング

1. 研究開始当初の背景

本研究「認知言語学に基づいた英語語彙・構文ネットワーク構築のための学習法開発」は次のような背景のもとに構想された。

(1) 社会的背景

英語教育の実質的な効果向上が求められている。大学などの高等教育機関においては、「グローバル人材」育成の必要性が叫ばれており、限られた時間で着実な効果を上げられる英語教材・学習法の開発が待ち望まれている。一方で、今日の一般的な教育課程において、学習者が習得している知識の量が必ずしも少ないわけではない。学習者が既に持っている言語知識を活かしながら、一方で根本的な再構築と補強を加え、真に有効なコミュニケーション手段として英語を使いこなせるようになるための仕組みが必要である。

(2) 理論的背景

近年、認知言語学(cognitive linguistics)と呼ばれる言語学の領域で積み重ねられてきた理論的知見を英語教育に応用することが盛んに行われている。認知言語学では、言語知識はスキーマ(schema)と呼ばれる抽象概念のネットワークとして構築される。スキーマを構成するのは形態素や語といった単位で表される概念だけではなく、句や文を作り出す抽象的な単位としての構文もまた複雑に絡み合ったネットワークを構成する。しかし認知言語学に基づく英語教育研究の多くは、個別的な語や表現の「イメージ」を記述するに留まっている。言語知識の断片を1つ1つ正確に描き出して示すことには確かに意味がある。しかしそれをそのまま教育の中で用いるならば、本来学習者がボトムアップ的に経験の中で自ら抽出すべき概念を、本人に成り代わって示すことになってしまう。本当に必要なことは、学習者自身が言語要素のスキーマを抽出し、それらをネットワーク化できることであり、限られた時間の中でこれを実現する方法論/学習法の開発が求められる。

(3) 技術的背景

今日、IT環境・技術の発達に伴い様々なデータやツールが数多く開発され、その一部は一般に広くオープンな形で配布されている。代表例としてはWikipediaの多言語データや授業運営に特化したオープンソース・ソフトウェアのMoodleなどがある。オープンなデータやツールには数多くの利点がある。例えば、1)経済的負担なく利用できること、2)多数の利用者や開発者が関わることで高い品質が確保できること、3)必要に応じて拡張が可能であることなどである。しかし、英語学習や英語教育に関連するデータやソフトウェア・ツールで、実際の現場のニーズに沿ったものは必ずしも多くない。今後の可能性を見据えたある種のモデルを示すことには大きな意味があると考えられる。

2. 研究の目的

上記のような背景のもとになされた本研究

の目的は、ある程度の基礎的知識を持った学習者が、語、句、文という複数のレベルで体系的に抽出、整理された用例に触れ、それらを内化させることで自ら言語知識のネットワークを構築できるような仕組みの構築である。研究開始当初に目指した内容の多くは3年の研究期間を経て達成されたが、一部については、詳細な検証の結果、修正された。

(1) 語彙リストの作成

学習者にとって「使いこなす」ことが求められる語彙リストの作成が1つ目の目的である。高校・大学レベルの学習者になると既知の語彙も豊富であるが、本研究では、単なる表面的な意味だけでなく、言語知識のネットワーク構築という観点から、他の語との関係が重要な語や、特徴的な用法の確かな理解が必要な語を抽出することを目指した。

(2) コロケーションおよび構文の抽出

それぞれの語彙項目は独立して機能しているわけではなく、他の語とのコロケーション(collocation)、すなわち色々な度合いでの結びつきと共に構文(construction)を構成している。構文は認知言語学にとって最も重要な概念の1つであり、やや極端な観点から述べるなら、文法とは様々な規模の構文知識に他ならない。したがって学習者の既存知識を再構築し、「使える」英語にするためには、重要な構文の数々を洗い出し、それらに習熟させることが不可欠である。

(3) 基本例文の作成

認知言語学の研究者が共有する言語観として言語の用例基盤(=使用依拠)モデル(usage-based model of language)がある。これによると、いわゆる文法知識を含む、あらゆる言語知識は実際に発話される言語表現が抽象化・一般化を経た結果としてのみ存在し得る。言い換えると、認知言語学では、従来の形式主義的な言語理論において想定されていたように、文法規則があらかじめ規定されており、それに基づいて語が組み合わさることで文が構成されるとは考えない。実際の句や文が先にあり、それらの意味や形式に対しある種のカテゴリー化のプロセスが働いた結果として文法が立ち上がると考えるのが認知言語学である。このような考え方に立つと、良質な文例を確保することが語彙や構文の学習にとって欠かせない。

(4) 学習効果の測定について

外国語としての英語学習の環境や方法の提案や改善にはある種の定量化が求められるため、当初本研究では語彙・構文リストを用いた学習を少人数のグループに課し、効果測定を行う予定であった。しかし小規模グループでの効果測定にとどまるのではなく、ウェブを通じて一般の学習者や指導者の使用に供する方向性に舵を切ることにした。前節の「研究の背景」を鑑み、社会のニーズに応えるべく成果を広く公開するための研究・開発に注力するという判断である。結果として学習効果の測定については課題として扱うに至った。

3. 研究の方法

ここでは前節で示した目的を達成すべく行った研究・開発の要点を示す。本研究では認知言語学の理論的枠組みを用いることを重要な特徴としており、成果物も認知言語学の概念や理念を反映させたものになる。そこで、研究の各段階で認知言語学上のどのような考え方を念頭に判断と作業を行ったかを特に明らかにしながら述べていきたい。

(1) TED コーパスと n-gram の作成

本研究では、多くの学習者にとって最適な語彙・構文リストを作成するため、まず、各種コーパスから頻度情報を得る作業を行った。当初、調査に用いたのは British National Corpus (BNC) および Corpus of Contemporary American English (COCA) である。しかし BNC や COCA における頻度情報は英語に関する基本的な事実を映し出すものとして重要であることは間違いないが、必ずしも英語学習に最適化されたものとは言えない。大規模な均衡コーパスを用いる限り、頻度リストは概ね似通ったものになる。あらゆる分野を総合して高頻度を示す語彙項目を学ぶことは重要ではあるものの、それだけでは単に既存の英語知識の大部分をなぞるに過ぎない。また、BNC や COCA などの大規模コーパスからは各語彙項目の実例を得ることはできても、様々なコンテキスト情報へのアクセスが困難である。

本研究で目指すのは受信・発信の両面を視野に入れた「実践的な」力につながる言語知識の構築である。そのために大規模な均衡コーパスを用いてデータの採取を行うことは必ずしも有効でないことがわかった。様々なジャンルの言語にはそれぞれの特徴がある。小説、ニュース原稿、政治演説、商業広告といったジャンルにはそれぞれの傾向があり、大規模均衡コーパスからのデータはそれらを（必要な調整を施しながら）バランスよく含むことを是としている。これまで学習者が学んできたのは、ある意味でそのようなデータに基づいた言語知識であるが、そこから一歩進めるためには、より実践的な言語運用の場面に特化したコーパスを用いて、大規模均衡コーパスでは頻度の面で上位に来ないものではあっても、実践的な英語運用の場面で有用と考えられる表現を多く取り入れた表現リストの作成が望ましい。

このような問題意識から、本研究の目的のために新たなコーパスとして TED Corpus を構築した。TED (<http://ted.com>) とは「広める価値のあるアイデア」を持った人物を集め、プレゼンテーションのカンファレンスを行っている団体であり、プレゼンテーションのデータは一定の条件のもとに広く利用可能である。また、データの利用を促進するために TED は約 2,000 のプレゼンテーションに関する詳細情報をコンピュータ上で取得するためのアプリケーション・プログラミング・インターフェイス (API) を公開している（ただ

し、現在はアクセス権の発行を停止している）。本研究ではこの API へのアクセス権を得て、合計すると約 500 万語となる英語プレゼンテーションのスク립トを取得、自然言語処理技術により統語解析を施した上で、様々な方法で語句や構文を検索できるシステムを作り上げた。TED Corpus Search Engine (TCSE) である。

TCSE には、単なる語彙の頻度情報だけではなく、いわゆる n-gram を取得できる機能を実装した。n-gram とはコーパスを構成するテキストの全体から n 個の連続する語を取り出したもので、TCSE では 2 から 4 までの n グラムを自在に検索できるようになっている。これを利用することで、語と語のコロケーションや、イディオムのなままとまりを考慮に入れた表現リスト作成が可能になった。

(2) 構文リストの作成

前述の n-gram 情報を活用するとともに、対象とする学習者にある程度の既存知識があることを想定し、複数の学習書や学習用語彙リストも参考にしながら、実践的な言語知識ネットワークの構築のために役立つ構文リストの作成を試みた。当初は語彙リストの作成と構文情報の抽出を別の取り組みとして計画していたが、理論的には両者の厳密な区別は恣意的なものであり、検討の結果、約 1,100 の構文パターンからなるリストとしてまとめることになった。なお、本リストは必ずしも n-gram の粗頻度に基づくものではなく、また今後の継続的な研究やコーパス自体の更新によって変更や拡張を加える可能性がある。いずれにしても、全表現が TED Corpus の複数のプレゼンテーションで使用されていることを 1 つの条件とし、前置詞や副詞などのいわゆる不変化詞に代表される部分要素を通じて個々の表現が互いに連関を持つような工夫を施して作成された。

(3) 例文セットの構築

加えて、約 1,100 の構文のそれぞれについて例文を 1 例ずつ作成した。もちろん構文の基本的な意味・用法が 1 つに定まることはほとんどない。それぞれの表現がある種の放射状カテゴリー (radial category) を構成しており、単一の例文でそれらを代表させることは認知言語学の基本的な考え方に反するとも言える。構文ごとに多数の例を列記することも可能ではあるが、あえて単一の例を付けた背景には 2 つの理由がある。まず、これらの例はあくまで典型例として、学習者の記憶に残りやすいことを目指したものである。認知言語学の考え方に基づくと、プロトタイプ (prototype) としての意味を身につけるには複数の事例に触れなければならない。しかし一方で、象徴的な少数の事例が典型 (exemplar) として、言語表現の理解や運用に大きく関わっている可能性も指摘されている。本研究における例文はまさにこの意味での「典型」として機能することを企図したものである。さらに、すべての表現リスト項目には作例に加

えて TCSE 用の検索文字列が付加されている。これを用いることで、約 2,000 の TED プレゼンテーションの中で実際に使われた事例が得られる。これが 2 つ目の理由である。これらの仕組みにより、全体として、典型例と具体事例を同時に提供する構成が可能になった。学習者は重要表現のそれぞれについて、シンプルで「教科書的」な作例を典型例として記憶にとどめた上で、コーパス中に見出される複数の事例に触れる。これは、本来、実際の言語経験の中で起こる語彙や構文との出会いを、オンラインのシステム上で擬似的に再現したものである。

(4) 音声認識による能動的学習活動の促進

前述の通り、本研究では当初、構築した構文リストおよびオンライン・システムを用いた学習プログラムを少数の学習者グループを対象にして教室内で実施し、効果を検証する予定であった。しかし試行の段階でいくつかの問題にぶつかることとなった。本研究が実現を目指す「学習者の既存の知識を最大限に活かし、より実践的な言語知識ネットワークを再構成する」ためのプロセスは多分に個人的であり、少数のグループとはいえ、画一的なプログラムで実現することはきわめて困難であることがわかった。一方で、本研究で構築した一連のシステムは、学習者が自らの知識の中で不足している部分を認識し、それを補完すべく自学自習を行う目的にきわめて適していることが明らかになった。そこで、ウェブ上で時間や場所の制約なしに自由にアクセスできる利点を活かし、成果を広く教育の現場や一般に提供するという観点から、教室内での実証実験ではなく、汎用性の高い学習環境の構築を目指すという判断に至った。そして、教室外でも能動的な学習活動を促進する方策の 1 つとして開発を進めたのが、コンピュータによる音声認識機能を利用して約 1,100 の構文リストの例文の発音・発話練習を可能にするシステム E-Speak である。

4. 研究成果

(1) 学習システムの仕様と特徴

ここでは前節で示した研究の成果として完成した学習システムの基本的仕様と特徴について述べる。多くの点が認知言語学的な必要性に基づいて発想されている。また、限られた期間ではあるが、大学における英語教育の現場での検証を行いつつ実装と調整を実施した。

① TED Corpus Search Engine

前述の通り、約 2,000 の英語プレゼンテーションのスク립トに統語解析を行ったデータが格納されており、すべての表現について表層形のほか、基本形 (= 動詞、名詞、形容詞の活用のない形) や品詞カテゴリーなどの情報を用いた検索が可能になっている。これにより、様々な構文や慣用句の事例を容易に得ることができる。

TED では英語プレゼンテーションのスクリ

プトの他、多言語による翻訳テキスト・データが利用可能となっている (いずれもボランティアの翻訳者たちがピア・レビューを行いながら作成している)。TCSE では日本語を含む 20 の言語による翻訳テキストも検索できるため、必要な表現を英語以外の言語から探すことも可能である。

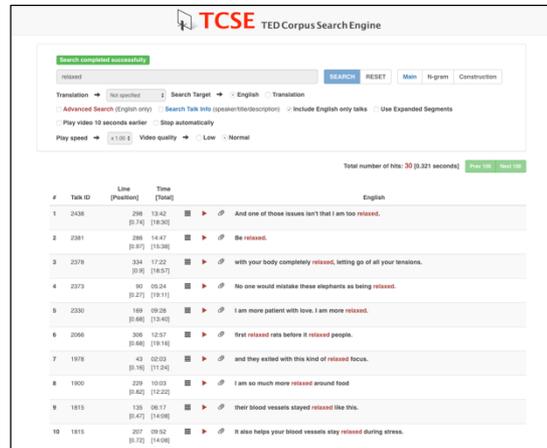


図 1: TCSE の検索画面

TCSE の機能の中で実践的な英語の学習にとりわけ役立つのは、検索結果として表示されるテキストについて、該当する場面の動画をピンポイントで再生できる機能である。認知言語学ではあらゆる表現の背後には発話のグラウンド (ground) があり、客観的内容がこれと概念的に結びつけられることではじめて言語表現として所定の効果を発揮すると考える。また、そのような実際の発話事象との接触を経て、命題的内容のみにとどまらない包括的な意味構造のスキーマが構築される。この考え方に立てば、TCSE は仮想的にはあるが、本来の形に近い言語事例の経験を可能にするツールであると言える。

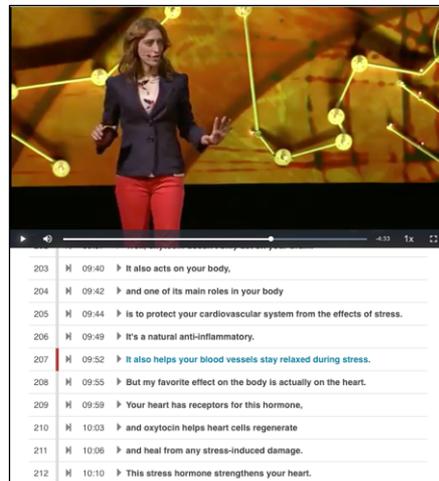


図 2: 構文事例のピンポイント再生

② 語彙・構文リスト

約 1,100 の語彙・構文リストは TCSE システム上で一覧・検索できる。部分検索が可能であるので、例えば特定の前置詞だけを含んだ構文を一覧することができる。これにより単にリストを定められた順に学習するのではなく、一定の意味的あるいは形式的な特徴を持った構文のグループを学習者自らが見出し、それらを集中的に学ぶなど、柔軟な利用が可能になっている。多くの言語表現が実際には単一の構文ではなく、大小の重なり合う構文群の実現形であるという認知言語学における構文文法 (construction grammar) の考え方に一致する仕様である。

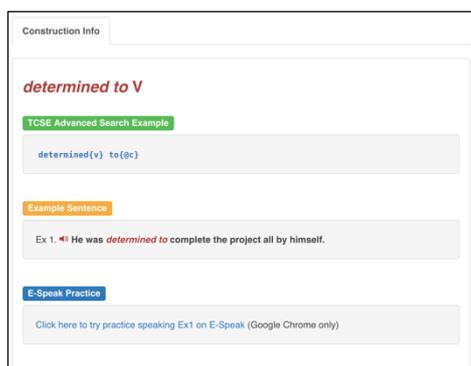


図 3：構文リスト項目の表示

構文リストの各項目は 3 種の情報と共に表示される。1 つは TCSE での検索キーであり、クリックすることで構文の実例を得ることができる。もう 1 つは作例であり、これには、合成音声による読み上げ機能が実装されている (対応ブラウザのみ)。そして 3 つ目は次に述べる E-Speak へのリンクである。

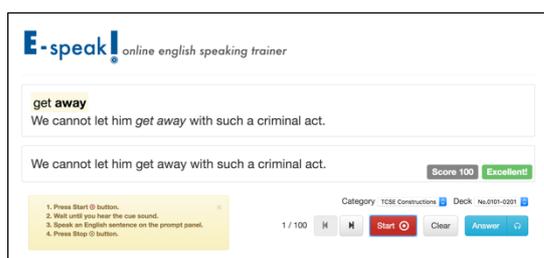


図 4：E-Speak による学習者の発話音声認識

③ 発話トレーニング・システム E-Speak

本研究で開発した成果物はすべてウェブ・ブラウザ上で利用可能となっている。ウェブ・ブラウザの標準仕様を決定する団体 W3C は 2012 年にウェブ・スピーチ API と呼ばれる音声認識のための機能を策定しており、現時点では Google Chrome において利用が可能になっている。これを用いて開発したのが E-Speak である。その主な目的は学習者の自学自習に主体的な発言的側面を加え、語彙や構文の定着度を高めることにある。このシステム上で、学習者は重要構文リストに対応する例文を発話し、正答との一致度に応じて表示されるス

コアを目安にしながら当該の表現の定着度を測ることができる。

(2) 本研究の成果もたらす影響

本研究が目指したのは限られた時間の中で実践的な英語力を育てる学習法の開発である。しかしながら研究の過程において、学習者のレベルが様々に異なるのに対し、画一的な学習法では十分な効果を見込めないことがわかった。一方で、画一的な方法を得ることは現実的でないとしても、様々なレベルの学習に対応する共通プラットフォームとしての学習システムを構想・実現することは可能ということが明らかになった。本研究の一連の成果物は、様々なレベルの学習に活用できる。また、TED のプレゼンテーション・スクリプトを中心とした豊富な言語資料と、一般公開されたウェブ上のインターフェイスは、学習や教育におけるベスト・プラクティスを互いに共有し、学習者間、あるいは指導者間の協力的な活動の中で、よりよい学習法を見出すことを促進すると考えられる。

本研究では、認知言語学が提示する方法論を最大限に用いた学習システムとして 1 つの形を示すことに成功した。ここまで示してきたように、次のような認知言語学上の考え方を重視した設計がなされている。「構文の概念の重視」、「典型例と個別事例の作用によるスキーマ的知識の獲得」、「発話のグラウンドに関する情報へのアクセス」などである。また、音声認識システムの活用により学習者の主体的な発話行為を促す点は、認知言語学における「身体化された知識」の考え方とも合致する。これらはいずれも、十分に準備された良質な言語習得環境では何らかの形で実現されているはずのものである。しかし言語学習の実際の場面には様々な制約があり、理想的な環境を得るのは簡単でない。本研究の成果は、認知言語学の理論的概念が示唆する「理想」を、オープンデータ、ウェブ技術、自然言語処理技術の助けを借りて可能な限り実現する試みであり、新たな取り組みに対しても 1 つのモデルを示すものとなるだろう。

(3) 今後の展望

本研究の成果に基づいた新たな展開には様々な可能性がある。ここでは教育的可能性、技術的可能性、言語学的可能性の 3 点について述べる。

① 教育的可能性

認知言語学の知見を言語教育に活かす取り組みは近年盛んに行われているが、その試みには本質的な困難が伴う。認知言語学な観点からすると、言語知識を構成するスキーマの概念の多くは経験の中で学習者自身が見出していかなければならない。一方で、言語教育の実際の場面では、実際の言語経験というプロセスを省きながら、エッセンスとしてのスキーマを身に付けさせることを目指している。例えば前置詞の意味構造をイラストで示し理解させようとする仕掛けは、そのような試みの典型である。つまり、本来ボトムアップであ

るべきプロセスを、トップダウンなプロセスに変換してしまっているのである。実際のところ、語学教育の営みにおいてこれは避けがたい。しかし、真の意味で認知言語学的なアプローチをとるならば、そこから一歩踏み込んで、学習者が上質かつ豊富な事例に触れ、構築されつつあるスキーマを補強する方法を検討する必要がある。またそのようなプロセスはできるだけ自発的・主体的なものであることが望ましい。認知言語学に基づく言語教育の方法論はこのような本質的な難しさに取り組むべきであり、本研究の成果をもとにそのような研究をさらに進めていきたい。

② 技術的可能性

認知言語学の考え方は多くの部分において非常に「直観的」である。経験された事例からスキーマを抽出し、それを元にして新たな場面での言語活動が展開される。ある意味で「当たり前」とさえ言えるプロセスであるが、このプロセスを（模擬的にであれ）再現することはこれまで困難であった。しかし、近年の技術の発展により、柔軟なテキスト検索システムの構築や動画・音声の自由自在な再生が可能になった。また高品質な音声合成や実用レベルの音声認識も利用可能である。本研究の試みを出発点に、今後さらに現れてくる様々な技術を言語教育に有効活用する手法の研究を進めていきたい。

③ 言語学的可能性

最後に、認知言語学という学問領域自体の発展と本研究との関連に基づく展望について述べる。認知言語学では豊富な事例に基づいた考察や統計的裏付けを得た議論を行うために、近年コーパスを用いた研究の重要性が高まっている。一方で、利用可能な大規模コーパスの多くは、問題となる言語現象について必ずしも十分なコンテキスト情報を与えてくれない。また、格納されたデータの形式や検索エンジンの制約によって、認知言語学が重視する「構文」レベルでの用例抽出が困難である。本研究の成果物であるシステムはデータの規模こそ限られているものの、いずれの面でも改善がなされている。今後は、認知言語学的な発想から生まれた本システムを用い、教育的、技術的な可能性を追求するだけでなく、認知言語学の理論的研究にも寄与するような研究を実施していきたい。

5. 主な発表論文等

[雑誌論文] (計1件)

- ① HASEBE, Yoichiro. Design and Implementation of an Online Corpus of Presentation Transcripts of TED Talks. *Procedia - Social and Behavioral Sciences*. 査読有, Volume 198, 174-182. 2015.

[学会発表] (計7件)

- ① 長谷部 陽一郎. オープンデータによる

英語構文事例検索システムの可能性: TED Corpus Search Engine を例として. 大学英語教育学会中部支部大会・シンポジウム「英語力向上のための多様なリソース活用の新展開」2016年6月4日. 愛知県立大学(長久手市).

- ② 長谷部 陽一郎. 事態把握の参照点設定と構文の選択について: コーパス調査に基づく試論. 成蹊CAPSプロジェクト第4回研究会. 2016年3月26日. 成蹊大学(武蔵野市).
- ③ HASEBE, Yoichiro. Constructions and long-tailed distribution of their collexemes: A Look at the relationship between low-frequency words and constructional prototypes. The 13th International Cognitive Linguistics Conference. 2015年7月20日. Northumbria University (英国・ニューカッスル).
- ④ HASEBE, Yoichiro. Design and implementation of an online corpus of presentation transcripts of TED Talks. The 7th International Conference on Corpus Linguistics. 2015年3月7日. University of Valladolid (スペイン・バリャドリッド).
- ⑤ 長谷部 陽一郎. TED Corpus Search Engine を用いたコンテキスト重視の英語構文学習. コーパス×学習者ラウンドテーブル. 2015年2月22日. 筑波大学東京キャンパス(東京都).
- ⑥ 長谷部 陽一郎. コーパスとコロケーション—認知言語学的視点から—. 同志社ことばの会. 2014年2月1日. 同志社大学(京都市).
- ⑦ 長谷部 陽一郎. コーパスから得られる頻度情報とプロトタイプ性について. シンポジウム「コーパスと言語学」. 2013年11月2日. 北海道大学(札幌市).

[図書] (計0件)

[産業財産権]

- 出願状況 (計0件)
- 取得状況 (計0件)

[その他]

ホームページ等

<http://yohasebe.com/tcse>

6. 研究組織

(1) 研究代表者

長谷部 陽一郎 (HASEBE, Yoichiro)

同志社大学・グローバル・コミュニケーション学部・准教授

研究者番号: 90353135