様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月 ２１ 日現在

機関番号：８２６２６
研究種目：基盤研究(C)（一般）
研究期間：2018～2022
課題番号：１８Ｋ１１４５６

研究課題名（和文）表現学習による語彙的変異の通言語的研究

研究課題名（英文）Cross-Linguistic Studies on Lexical Differences based on Representation Learning

研究代表者
高村　大也（Takamura, Hiroya）

国立研究開発法人産業技術総合研究所・情報・人間工学領域・研究チーム長

研究者番号：８０３６１７７３

交付決定額（研究期間全体）：（直接経費）　　3,300,000円

研究成果の概要（和文）：ロマンス語の同源語を対象に、単語の意味的差異と、頻度や多義性など6つの変数との間の統計的関係を調査した。意味的差異の程度は、単語分散表現の余弦距離を使って定量化した。回帰分析を行い、意味的差異に対し、頻度は負の影響が、多義性は正の影響があることを実証した。さらに、形態的に複雑な語根は意味変化が起こりにくいこと、長い期間使用されてきた同源語は意味変化を起こしやすいことを明らかにした。また、”better off”という表現の新しい用法が定着した経緯に関する仮説を、単語分散表現を用いて検証した。また、社会的な違いによる語彙的変異として、母語話者と非母語話者による違いに着目し調査を行った。

研究成果の学術的意義や社会的意義
単語分散表現を含む深層学習技術は、言語研究における新たな道具であり、それを実証する成果が得られている。これまで変化検出の研究が多かった中で、語彙的変異の要因を探った点で学術的意義が大きい。また、"better off"に関する研究では、言語学で考えられた仮説を検証しており、自然言語処理技術の言語学への貢献の形として、良い例となるだろう。

研究成果の概要（英文）：We investigated the statistical relationship between semantic difference in Roman cognates and six variables including frequency and polysemy. The degree of semantic difference was quantified using the cosine distance of the distributed representations of words. We conducted regression analysis and demonstrated that frequency is negatively correlated with semantic difference, while polysemy is positively correlated with semantic difference. We also found that morphologically complex word roots are less likely to undergo semantic change, while cognates that have been in use for a long time are more likely to undergo semantic change. We also examined how the new usage of "better off" came to be established. In addition, we investigated the lexical variation between writings by native speakers and non-native speakers.

研究分野：自然言語処理

キーワード：語彙的変異　分散表現　深層学習　意味変化

１．研究開始当初の背景

英単語 *nice* が 1300 年代は「*愚かな*」の意味であったように、単語の意味は変化する。通時的な変化だけでなく、地理的要因での違いもあり、コミュニティで特殊な意味を持つこともある。このような**語彙的変異**は古くから研究対象であったが、網羅的に分析するための技術は確立されていない。一方、単語を含む言語表現を数学的に表現する手法が、表現学習と呼ばれる枠組みの中で発展してきている。特に、単語の**分散表現**と呼ばれるベクトル表現は、自然言語処理で非常によく使われている。 分散表現は、単語の意味の差異を測る手段として性能が高く、語彙的変異の研究の新たな有効手段となることが期待される。実際、通時的な意味変化を分析した研究や、地理的な影響による意味変化を扱った研究、外来語の意味変化の研究などがある。しかしこれらはまだ非常に初歩的な段階にあり、研究対象が英語だけである、地理的な近さなどの情報がモデルに取り入れられていない、時間と空間が同時に扱われていない、変異の種類が扱われてないなどの問題点がある。

２．研究の目的

本研究では、申請者らが分散表現を用いて日本語外来語の意味変化を分析した研究をさらに進め、変異の検出、変異の種類の分類などの分析を行う。また、そのための技術開発を行う。分析対象は、同源語の語彙的変異や、地理的あるいは社会的な違いによる語彙的変異を含む。

３．研究の方法

時代や社会的コミュニティなどが異なる複数のコーパスを用い、それぞれで単語分散表現を計算し、同じ空間に写像する。単語分散表現を比較することで、意味変化が起こった単語を検出する。また、変化度合いと様々な特徴量との関係を見ることで、どのような特徴を持つ単語が変化しやすいか、どういった変化が起こると考えられるか、などを考察する。また、事例ごとに分散表現を計算し、その分布を見ることで、いろいろな用法の経時的変化を調査する。

４．研究成果

ロマンス語の同源語を対象に、単語の意味的差異と、頻度や多義性など 6 つの変数との間の統計的関係を調査した。同源語とは、共通の語源から派生した単語である。語源を共有しているにもかかわらず、いくつかの同源語のペアは意味的な異なっている。意味的差異の程度は、同源語の対応する単語分散表現の余弦距離を使って定量化した。先行研究では、頻度と多義性は意味的差異と相関があると報告されているが、様々な方法論上の欠陥があるため、修正の上、再調査をすることが必要である。本研究では、意味が変わらない同源語および意味的差異がある同源語を検出する実験、および意味的変化を予測する回帰分析を行った。これにより、実際に意味的が変わらない同源語、意味的差異がある同源語を検出した。例を Table 1 に挙げる(表は、研究業績である Kawasaki らの COLING での発表論文からの抜粋である)。

French	Spanish	Sim.
construire “to construct”	constuir “id.”	0.87
provoquer “to provoke”	provocar “id.”	0.87
évêque “bishop”	obispo “id.”	0.86
détruire “to destroy”	destruir “id.”	0.86
féminin “feminine”	femenino “id.”	0.86
avoyer “to set saw”	aviar “to prepare”	0.05
atteindre “to reach”	atañer “to pertain”	0.05
mener “to take”	menar “to turn”	0.04
saison “season”	sazón “seasoning”	0.02
maire “mayor”	mayor “bigger, older”	0.00

Table 1: Most (upper half) and least (lower half) similar five French–Spanish cognate pairs.

また、意味的差異に対し、頻度は負の影響が、多義性は正の影響があることを実証した。さらに、形態的に複雑な語根は意味転換に強いこと、より長い期間使用されてきた同源語はより大きな意味変化を起こしやすいことを明らかにした。

また、単語や言語表現の用法の変化および新しい用法が定着した経緯を調査する手法を提案した。ケーススタディとして、”better off”という表現の新しい用法が定着した経緯に関する仮説を検証した．具体的には，コーパス中の用例をクラスタリングすることで用法変化を定量化し，言語学的な観点から提示された仮説に一致するかどうかを検証する．その結果は，従来の仮説を大

枠で支持するものの，派生するいくつかの可能性を示唆することがわかった．また，分類結果に基づき，関連した新たな研究の方向性を示した。

**　また、社会的な違いによる語彙的変異として、母語話者と非母語話者による違いに着目し、これについても調査を行った。**

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　0件／うちオープンアクセス　1件）

１．著者名 Aida Taichi、Komachi Mamoru、Ogiso Toshinobu、Takamura Hiroya、Mochihashi Daichi	４．巻 30
２．論文標題 異なる時期での意味の違いを捉える単語分散表現の結合学習	５．発行年 2023年
３．雑誌名 Journal of Natural Language Processing	６．最初と最後の頁 275～303
掲載論文のDOI（デジタルオブジェクト識別子） 10.5715/jnlp.30.275	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計7件（うち招待講演　0件／うち国際学会　0件）

１．発表者名 川崎義史，Maelys Salingre（東大），Marzena Karpinska（University of Massachusetts Amherst），高村大也（産総研），永田亮（甲南大）
２．発表標題 分散表現を用いたロマンス語同源語動詞の意味変化の分析
３．学会等名 言語処理学会第28回年次大会
４．発表年 2022年

１．発表者名 永田亮（甲南大），大谷直輝（東京外大），高村大也（産総研），川崎義史（東大）
２．発表標題 言語処理的アプローチによる better off 構文の定着過程の説明
３．学会等名 言語処理学会第28回年次大会
４．発表年 2022年

１．発表者名 永田亮（甲南大），高村大也（産総研）
２．発表標題 文法誤り訂正への訂正重要度の導入
３．学会等名 言語処理学会第28回年次大会
４．発表年 2022年

１．発表者名
相田太一、小町守、小木曽智信、高村大也、持橋大地

２．発表標題
通時的な単語の意味変化を捉える単語分散表現の同時学習

３．学会等名
言語処理学会 第27回年次大会 発表論文集

４．発表年
2021年

１．発表者名
相田太一 小町守 小木曽智信 高村大也 坂田綾香 小山慎介 持橋大地

２．発表標題
単語分散表現の結合学習による単語の意味の通時的変化の分析

３．学会等名
言語処理学会 第26回年次大会 発表論文集

４．発表年
2020年

１．発表者名
井上誠一，小町守，小木曽智信，高村大也，持橋大地

２．発表標題
Infinite SCAN: 単語の意味変化と語義数の同時推定

３．学会等名
信学技報

４．発表年
2022年

１．発表者名
永田亮、高村大也

２．発表標題
自然言語処理を利用した実証的な認知言語学の可能性

３．学会等名
日本英文学会全国大会 シンポジア第11部門

４．発表年
2022年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

「分散表現を用いたロマンス語同源語動詞の意味変化の分析」は言語処理学会第28回年次大会にて委員特別賞を受賞

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	永田　亮 (Nagata Ryo) (10403312)	甲南大学・知能情報学部・准教授 (34506)	
研究分担者	川崎　義史 (Kawasaki Yoshifumi) (40794756)	東京大学・大学院総合文化研究科・准教授 (12601)	

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

8．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
米国	University of Massachusetts Amherst			