

令和 5 年 10 月 24 日現在

機関番号：27101

研究種目：基盤研究(C)（一般）

研究期間：2020～2022

課題番号：20K00667

研究課題名（和文）百科事典的知識の動的変化モデルの構築

研究課題名（英文）Constructing a dynamic model of encyclopedic knowledge: A corpus-based approach

研究代表者

木山 直毅 (Kiyama, Naoki)

北九州市立大学・基盤教育センター・准教授

研究者番号：20803894

交付決定額（研究期間全体）：（直接経費） 2,300,000円

研究成果の概要（和文）：本研究では、語の意味が、文脈の影響をどのように受けているのかを大規模なデータ（コーパス）を用いて調査を行った。そして、語彙意味論の研究で、トピックモデルの手法が有効であることを論じた。具体的には、(1) 語の意味が、話されている内容（トピック）の影響を受けるのかということと、(2) 語の意味の使用が社会的関心の影響を受けるのかということ进行调查した。また、これらの調査から、意味変異（semantic variation）が、社会的側面と強く結びついているという示唆を得ることができた。

研究成果の学術的意義や社会的意義

語の意味理解の研究において、トピックが考察されてこなかった。その要因の一つとして、トピックは非常に曖昧とした概念だからだと考えられる。本研究では、自然言語処理の分野で近年用いられることの多いトピックモデルの手法を援用することで、語の意味とトピックとの間を統計的に規定する方法を提案した。この考えは、長文読解における多義性指導に役立つと考えられる。そのため、英語教育への応用可能性も示唆される。また、認知言語学が、他の分野と相性が良いことが示唆され、今後の学際的な協力関係を築くきっかけにもなるだろう。

研究成果の概要（英文）：In this study, we investigated how the meaning of words is influenced by context using a large corpus. Then, we argued that the biterm topic model, one of the topic modelling methods, is effective in the study of lexical semantics. Specifically, we investigated (1) whether word meanings are influenced by the content, or what is being talked about, and (2) whether the use of word meanings is influenced by social interest. These investigations suggest that semantic variation is strongly related to social aspects.

研究分野：認知言語学

キーワード：認知言語学 テキストマイニング 社会的転回 ジャンル研究

1. 研究開始当初の背景

従来より認知言語学では、「言葉の解釈」には問題となる語(句)について、文脈より適切な背景知識(もしくは百科事典的知識)の喚起が必要だと考えられてきた(Lakoff 1987, Langacker 1987, Fillmore 1987, Haiman, 1980; 他; なお、談話分析や語用論の多くの研究でも背景知識の重要性は長く論じられてきた)。ここで言う「適切な背景知識」とは文化モデルに相当するものであり(Kecskes 2013)、換言すると、それは国や地域、さらには時代によって絶え間なく変化するものである。このことは、背景知識に基づく言語研究にとって、話者の知識の変化を捉える視点を持つ必要があることを含意する。しかし、従来の研究では通時的観点から百科事典的知識の動的な側面を論じたものはほとんどなかった(Peirsman et al., 2010 は重要な例外)。また、背景知識が文化に根付くものであるならば、同一言語であっても、言語が使用されるテキストジャンルが異なれば使用される背景知識が異なってくると考えられる。そのため、文脈が百科事典的知識の喚起にどのように影響するのか、また、時代の変化やジャンルの差が、百科事典的知識にどのように影響を与えるのかの解明は、意味を概念化として捉える理論にとって必要だと考えられる。

本研究では、英語の動詞 run を取り上げて研究する。この動詞は非常に多くの意味を持つ動詞であり(多義語)、Merriam Webster に記載されている意味は 30 を超える。そのため、認知言語学においても多くの意味分析がなされてきた(e.g., Glynn, 2014; Gries, 2006; Langacker, 1988)。たしかに、先行研究では、英語の動詞 run は様々な意味を持つことが論じられてきたが、彼らの意味記述においては、run が使用される際の主語や目的語に着目しており、語が使用される文脈・談話が何についての話なのか、あるいはトピックが語の意味解釈に与える影響については、あまり研究されてこなかった。しかし、たとえば新聞のスポーツ欄において run が用いられれば、run は「人の高速移動(足をはやく動かすことによる前進移動)」の意味で使用される可能性が、他の意味で使用される可能性よりも高いことが予想されよう。その場合には、「選挙での候補者に名を連ねる」という意味で用いられている可能性が高いことが予想されるであろう(Deignan, 2005; Taylor, 2003 参照)。

上述した意味の違いに加えて、意義の違いもトピックと関わりがあると考えられる。たとえば、スポーツ競技内で走っている様子と、泥棒が警察から逃げる目的で走っている様子というのは、事態に関わる参加者に違いがある。具体的には、泥棒が逃走する場合、必ず「誰から逃げるのか」という人物(あるいは組織)が必要であるが、スポーツにおいては、その参加者は不要であろう。百科事典的意味論において、参加者に差があるということは、使用される構文も異なると想定される(Fillmore, 1987; Fillmore & Atkins, 1992 etc.)。以上の直感は、英語の動詞の解釈においては、言葉が用いられる談話、あるいはトピックが重要な役割を持つことを示唆している。

本研究における百科事典的知識とトピックは、ほぼ同じものとみなすことができる。たとえば、上述した run を例にとってみると、スポーツに関する知識が、意味の特定を促進するための背景知識として働いていると考えられる。同じように、選挙に関するトピックが話題に上がっている場合、選挙に関する背景知識が語の意味の特定に繋がっていると考えられるからである。

このような直感から、本研究では、(i) トピックが百科事典的知識と強く結びつくとするならば、語の意味解釈や構文選択は、語や構文が使用されるトピックを解析することで明らかにできる、(ii) トピックが意味解釈に影響を与えるならば、テキストジャンルによって百科事典的知識は異なる、(iii) トピックの出現率変化によって、百科事典的知識の変化が観察できる、という 3 つの仮説を立て、百科事典的知識の通時変化と社会性に関する動的側面を、コーパスを用いて調査を行う。

トピックが語彙の意味解釈に影響を与える可能性については、申請者が過去に行った予備調査で確認された。その研究では、米国大統領の一般教書演説をコーパスとして用いており、ある語が使用されるトピックは実に多様であり、米国大統領の演説を理解する上で、様々な背景知識が関わっていることが明らかになった。また、当該調査により、分析方法としてトピックモデル(e.g., Blei et al., 2003; Yan et al., 2013)を援用することの有効性も確認された。

そこで本研究では、申請者の予備調査結果を学術的背景とし、言語に関わる背景知識とその変化、そして社会性という変数を用いて英語の多義語の調査を行い、百科事典的知識・意味論の動的側面を研究する。

本研究では、研究課題の核心をなす学術的問いとして、「ヒトが持つ知識はどのように変化し、それは何が要因となって変わるのだろうか」を設定し、それを解決する目的として、英語の多義語とトピックの関係を明らかにする。

2. 研究の目的

本研究は、上述の学術的問いに答えるため、英語の多義語とトピックの関係を明らかにすることを目的とする。具体的には、英語の動詞 run に関して、テキストジャンル間の違い、通時性、構文の選択などの変数とトピックの関係性の解明を目指す。

本研究の独自性は、認知言語学における百科事典的意味論に対して、テキストジャンル(e.g.,

Biber & Conrad, 2009; Conrad, 2015; Seoane & Biber, 2021) や通時性 (e.g., Gries & Hilpert, 2008; Hilpert, 2008; 2014; Ohashi; Perek, 2018) といった観点を取り入れ、トピックモデルを援用して分析点にある。これらの側面を取り入れた考察は、これまでの多義語研究には見られなかった。

本研究は、主として認知言語学の立場に立脚する研究である。近年、認知言語学は、文法構造や意味、語法といった、伝統的に言語学で扱われてきた言語構造へのアプローチに加え、言語習得 (Tomasello, 2003; 児玉・野澤, 2009) や語学教育 (長, 2016; Littlemore, 2009; Tyler, 2012) といった関連分野への応用も見られ、注目を浴びている。また、テキストジャンル間の比較を行うという点については、認知言語学と社会言語学の融合を目指す認知社会言語学の流れを汲む研究でもある (e.g., Harder, 2010; Kristiansen & Dirven, 2008; Pütz et al., 2014)。以上のように、本研究の知見は、伝統的な言語学の分野に加え、関連領域への影響力も秘めており、創造的な研究であると考えられる。

3. 研究の方法

本研究は、大規模言語データベース (コーパス) を用いた研究である。本研究が利用したコーパスは、次のとおりである。

- (i) Corpus of Contemporary American English (COCA): 1990年から2019年までのアメリカ英語を約10億語収録したコーパス
- (ii) News on the Web (NOW): 2010年から現在までの間に公開された英字オンラインニュース (世界20カ国分) を収集したコーパスで、約20億語を収録している。本コーパスは、1日にごとにデータを収集している。

本研究は、上記のコーパスに対して、次の統計手法を用いて解析を行った。

- (i) 潜在的ディリクレ配分法 (Latent Dirichlet allocation, Blei et al., 2003), バイタームトピックモデル (Biterm topic model, Yan et al., 2013) トピック計算に利用した。
- (ii) 多次元尺度構成法
 - (i) で得られたトピックをまとめるために利用した。
- (iii) 可変性に基づく近隣クラスタリング (Variability-based Neighbor Clustering, Gries & Hilpert, 2008) トピックの出現率の変化を調査する際、時系列に並んだデータがどのようにまとまるのかを考察する際に利用した。

4. 研究成果

【COCAとNOWの比較】COCAを利用する際は、データ量の都合から、フィクションデータに限定し、テキストジャンルの比較を行った。これにより、*run one's finger* のような「身体部位を動かす」の意味は、ニュースでの利用があまり見られないことがわかった。また、「人の高速移動」の意味は両方のジャンルに現れる意味であったが、フィクションでは *run from X to Y* のような経路を表す表現が多かった一方で、ニュースでは *run X miles/a marathon* のような出場科目等に関連する表現が目立った。これらの結果は、語の意味研究におけるテキストジャンルの差を考慮した研究の重要性を示唆している。

【時系列変化】NOWコーパスを利用して、意味使用の変化を考察した。図1は、動詞runの出現頻度を時系列に並べたものである。頻度に変化はあるものの、頻度が上がり続けている傾向が確認できる。もし、様々な意味の利用が通時的に安定したものであれば、図1のような傾向を示すはずである。しかし、実際は図2で示すように、意味の使用確率は一樣ではないことが明らかになった。図2の結果は、時の情勢に強く影響を受けていることを論じ、社会の中の意味 (meaning-in-society, Harder, 2010: 3) を考慮に入れた研究を行うことの重要性が示唆された。

図1 runの出現頻度変化

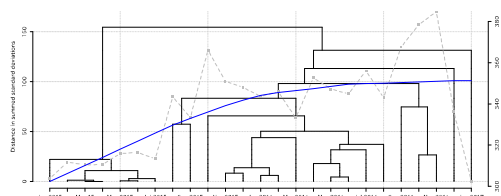
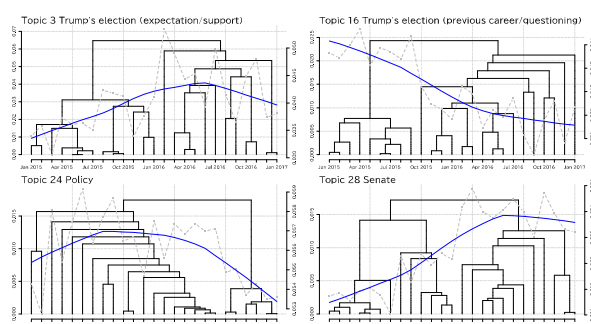


図2 トピックごとの出現率変化



5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 3件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 木山直毅・渋谷良方	4. 巻 0
2. 論文標題 動詞の意味はトピックから推測できるのか 英語の動詞 run を例に	5. 発行年 2021年
3. 雑誌名 英語コーパス学会大会予稿集2021	6. 最初と最後の頁 103-108
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 木山直毅	4. 巻 0
2. 論文標題 直接話法におけるaskと疑問符	5. 発行年 2021年
3. 雑誌名 語法と理論との接続をめざして	6. 最初と最後の頁 321-341
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Kiyama Naoki and Yoshikata Shibuya	4. 巻 0
2. 論文標題 Applying topic models to study polysemy: The case of the noun streams	5. 発行年 2021年
3. 雑誌名 日本認知言語学会論文集	6. 最初と最後の頁 291-303
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 木山直毅・渋谷良方	4. 巻 30
2. 論文標題 トピックモデルによる多義性研究：英語動詞run を例に	5. 発行年 2023年
3. 雑誌名 英語コーパス研究	6. 最初と最後の頁 47-70
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Naoki Kiyama and Yoshikata Shibuya	4. 巻 -
2. 論文標題 A Topic-Based Diachronic Account of the Polysemy of the English Verb Run	5. 発行年 2023年
3. 雑誌名 Research in Language	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Kiyama Naoki	4. 巻 -
2. 論文標題 Lumping genres together may "run" into a problem	5. 発行年 2023年
3. 雑誌名 日本認知言語学会論文集	6. 最初と最後の頁 500-505
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計5件 (うち招待講演 0件 / うち国際学会 1件)

1. 発表者名 木山直毅・渋谷良方
2. 発表標題 動詞の意味はトピックから推測できるのか --英語の動詞 run を例に--
3. 学会等名 英語コーパス学会第47回大会
4. 発表年 2021年

1. 発表者名 木山直毅・渋谷良方
2. 発表標題 文脈の定量化と語彙意味論-英語の多義語をトピックから考察する-
3. 学会等名 福岡認知言語学会第43回大会
4. 発表年 2021年

1. 発表者名 Kiyama Naoki and Yoshikata Shibuya
2. 発表標題 Applying topic models to study polysemy: The case of the noun streams
3. 学会等名 日本認知言語学会
4. 発表年 2020年

1. 発表者名 Naoki Kiyama and Yoshikata Shibuya
2. 発表標題 Semantic change and changes in social structure
3. 学会等名 Cognitive Linguistics in the Year 2022 (国際学会)
4. 発表年 2022年

1. 発表者名 Naoki Kiyama
2. 発表標題 Lumping genres together may "run" into a problem
3. 学会等名 日本認知言語学会
4. 発表年 2022年

〔図書〕 計1件

1. 著者名 木山直毅	4. 発行年 2021年
2. 出版社 開拓社	5. 総ページ数 13
3. 書名 意味論・語用論とコーパスのインターフェイス: トピックモデルを語彙意味論に応用する	

〔産業財産権〕

〔その他〕

-

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	渋谷 良方 (Shibuya Yoshikata) (70450690)	金沢大学・歴史言語文化学系・准教授 (13301)	

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