

## 「論文」

### トピックモデルによる多義性研究：英語動詞 *run* を例に\*

木山 直毅・渋谷 良方

#### Abstract

Research on polysemy takes a variety of approaches, including studies that are of a more introspective nature grounded in theoretical linguistics (e.g., Pustejovsky, 1995; Lakoff, 1987; Ravin and Leacock, 2002; Tyler & Evans, 2003), as well as those that are more empirically oriented, conducted from the perspectives of corpus linguistics (e.g., Sinclair, 1991; Tognini-Bonelli, 2001) and psycholinguistics (e.g., Kishner & Gibbs, 1996; Gibbs & Matlock, 2001) (for a comprehensive overview of polysemy research, see Nerlich et al., 2003 and Gries, 2015). A common perception in polysemy studies is that the meaning of a word varies with ‘context’. The same is true of research in pragmatics. Although context is a useful and important concept, it is also an ambiguous one. This is because there are so many variables involved in it; in addition to co-occurring words, social(-interactional) variables such as the flow of conversation, the context of utterance, and even the age and gender of the speaker can also be part of a context, depending on the theoretical assumption of the study (Stefanowitsch & Gries, 2008, p. 135). The present study reinterprets the notion of context from the viewpoint of topic modeling (e.g., Blei et al., 2003), an NLP method of classifying texts by analyzing the type of topic they are related to, on the basis of the distribution of words. By applying topic modeling, this study explores the polysemy of the English verb *run*. Our chosen topic modeling technique is the one called the Biterm topic model (BTM; e.g., Yan et al., 2013). Based on the analysis of a number of automatically generated topics and the corresponding meanings of *run*, we argue that the context in which the verb is used, or more specifically, the topic in the text, plays a crucial role in identifying the meaning.

## 1. はじめに

語は一般的に複数の意味を持つ。<sup>1</sup> 例えば、英語の動詞 run を Merriam-Webster (n.d.) で調べると、自動詞用法と他動詞用法を合わせて 30 個に及ぶ意味が見つかる。語が複数の意味を持つ現象は多義性 (polysemy) と呼ばれており、この現象についてはこれまで様々な研究が行われてきた。例えば、形式主義の観点からは生成語彙論 (Generative Lexicon) による分析 (e.g., Pustejovsky, 1995) があり、認知言語学の枠組みからは Lakoff (1987), Ravin and Leacock (2002), Tyler and Evans (2003), Glynn (2016) などがある (認知言語学における、多義性の研究の要約は、Glynn, 2014b 参照)。また、コーパス言語学の分野では Sinclair (1991), Tognini-Bonelli (2001) などがあり、心理言語学においては Gibbs and Matlock (2001), Kishner and Gibbs (1996), Ramsey (2022) などがあげられる (多義性についての包括的な説明は Nerlich et al., 2003 や Gries, 2015 参照)。

上述の通り、語の多義性についてはこれまで様々な方面から多くの研究が行われてきたが、先行研究には共通する認識が見られる。それは、文脈によって意味が変化するという見方である。(なお、文脈によって意味が変化するという考えは多義性研究に特化した認識ではない。例えば、語用論研究では、語 (あるいは言語表現) の意味を特定するのに文脈が必要であるという認識のもと、意味が理解されるメカニズムが論じられてきた (Bublitz & Norrick, 2011)。

意味や理解の研究では「文脈」という用語が広く使用されてきたが、文脈という概念について一枚岩的な定義は与えられてこなかった。例えば、文脈は共起語や会話の流れを指すこともあるし、発話状況や、さらには年齢や性といった社会的変数を文脈と呼ぶこともある。すなわち、文脈という用語が指す内容は研究目的によって変わりうるのである (Stefanowitsch & Gries, 2008, p. 135)。

このように、「文脈」とは本来非常に多くの変数 (要因) を含む概念である。意味の研究において文脈に説明を求めるアプローチでは、文脈という概念により、具体的にどの要因を考察対象としているのかを明確にするべきであろう。さらに、実証的方法で多義性を扱いたいなら、客観的に規定可能な方法に基づいて文脈を定義すべきである。このような状況を背景として、本研究では、語が用いられる内容、より具体的には、トピックモデルにおけるトピックの観点から (e.g., Blei et al., 2003) 文脈を定量化し、語の多義性を記述する方法を提案し、その有用性を論じる。

本稿の構成は次のとおりである。まず第2節では、本稿が想定する理論的背景と、多義性に関する先行研究を概観する。続く第3節では、本稿のリサーチデザインを紹介する。第4節では2つのコーパスを用いた事例研究を紹介し、トピックが英語の動詞の意味とどのような関わりにあるのかを論じる。第5節では、第4節で得られた結果が言語研究にとってどのような理論的含意を持つのかを論じる。第5節ではまた、本稿が使用する手法と先行研究の手法の比較も行う。

## 2. 背景と先行研究

### 2.1. 百科事典的意味論

伝統的な意味論（e.g., 形式意味論）では、意味は言語表現と参照物の間の対応、すなわち外延によって規定可能であるという想定の下に研究が行われている（cf., Cann, 1993）。このアプローチとは対照的な意味観を持つのが認知言語学に見られる百科事典的知識に基づくアプローチである。このアプローチでは、人が持つ世界に関わる知識との関連で、語の意味が捉えられている（cf., Haiman, 1980; Fillmore, 1982; 1985）。ここでは一例として、Taylor（2002）による *photograph* という英語名詞の説明を紹介する。Taylor は、*photograph* には少なくとも下記の4つの背景知識（ドメイン）が関わると述べている（Taylor, 2002, p. 442）。

- (i) 写真は、情景を描写したものである。
- (ii) 写真は、特殊な装置や工程、撮影者の腕前といった様々な技術を通して制作される。
- (iii) 写真は、通常、紙に現像された映像であるが、映像は硝子板や電子データのような媒体としても存在する。さらに、同じ映像を他の媒体へと複製することが可能である。
- (iv) 写真は、社会的慣習と密接に結びついたものである。ある人は人生の記念において写真を撮り、ある人は写真で生計を立て、ある人は写真を身分証として使用する。

(1) の事例は、Taylor が上記 (i)～(iv) のドメインが関わる事例としてあげたものである（事例は Taylor, 2002, p. 442 から）。

- (1) a. The photograph is torn.  
 b. This is a photograph of me at age 10.  
 c. This photograph has been re-touched.  
 d. The photograph was awarded a prize.  
 e. I'll send you the photograph as an electronic attachment.

Taylor の説明は以下の通りである (pp. 442f)。 (1a) では、(i) の知識が前景化される一方で、(ii) と (iv) の知識は背景化されている。(1b) では個人の幼少期が描写されているが、ここでは (iv) の社会文化的慣習が前景化されている。(1c) では修正 (re-touch) を可能にするテクノロジー、すなわち (ii) の前半部に関する知識が前景化されている。対照的に、(1d) では撮影者の感性や腕前が関わっており、(ii) の後半部分が前景化されている。(1e) では紙媒体以外の方法で保存できるという (iii) の側面が前景化されている。

上述の 4 つの知識のどれが *photograph* の意味と関係する背景知識なのだろうか。伝統的な意味論では、いわゆる語用論の意味や知識と語の意味は厳密に区別されるため、*photograph* の中心的な知識であると考えられる (i) 以外は *photograph* の意味とは無関係だとみなされるかもしれない。一方、百科事典の意味論では、世界に関する知識そのものが語の意味の一部だと考えられることになる (Croft, 1993)。すなわち、百科事典の意味論では、4 つの *photograph* の理解に関わる背景知識のうち、どれがこの語の言語内的意味で、どれが言語外的意味であるのかを明確に区別することができないと考え、言語に関わる世界の知識が語の意味であると考え (Taylor, 2012)。

## 2.2. コーパスを用いた多義語の先行研究と本稿の仮説

百科事典の意味観に基づく研究は主に認知言語学の枠組みに依拠する形で行われており、ここではコーパスを通じて実際の言語使用を調べる研究が積極的になされてきた (e.g., Akita, 2012; Glynn, 2009; 2016; Glynn & Robinson, 2014; Heylen et al., 2015; Hoffmann et al., 2019; Lederer, 2019; Schönefeld, 2013; Stefanowitch & Gries, 2003)。それらの中で、本稿に特に関連する分析が Gries (2006) や Divjak and Gries (2006) で提案された *behavioral profiles* (以下、BP 分析) である。BP 分析では、調査対象となるターゲット語が現れる統語環境に加え、それに伴う形態素、文法的数、アスペクト、意味カテゴリといった極めて幅広い特徴のコーディングがコーパスから収集されたデータに対してなさ

れ、階層的クラスター分析や多次元尺度法といった多変量解析を通じてターゲット語に関する様々な特徴が捉えられる。

BP 分析では、これまで get や run などの高度に多義的な英語動詞の意味記述が行われてきた (e.g., Berez & Gries, 2008; Divjak & Gries, 2008; Glynn, 2014a; Gries, 2006; 2010; Gries & Divjak, 2009; Gries & Otani, 2010; Jansegers & Gries, 2017)。BP 分析の結果は、語の意味を説明するためには語が現れる様々な環境を考慮に入れることが極めて重要であることを明らかにするものであった。上述のように、BP 分析では、数多くの要因についてコーディングがなされる。しかし、同分析法ではコーディングの対象外とされる要因もある。その一つが本研究が注目するトピックである。

本稿が論じる「トピック」とは、語の分布で表現される文書的话题を指す。本研究の仮説は、人が多義語を理解するにあたって、テキスト中で話題になっていること、すなわちトピックの観点から理解しているというものである。この仮説は次のような直感的理由に基づいている。たとえば、陸上の大会という特定のトピックに関する新聞記事の中で英語の動詞 run が用いられる場合について考えてみよう。そこでは、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）(fast pedestrian motion)」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。一方、選挙に関するトピックで run が現れた場合はどうであろうか。おそらく、その場合には、「選挙での候補者に名を連ねる (to become a candidate)」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。本稿ではこのような考えに基づき、動詞の意味は語が用いられるトピックとの関係で決まるという仮説を立てている。<sup>2</sup>

では、動詞の意味と、語が用いられるトピックの関係は、どのようなものだろうか。2.1 節で論じたように、百科事典の意味論では、語の意味は、背景知識（ドメイン）の中で指定されると考えられている。たとえば、run の「人の高速移動」の意味を理解する際に、人は関連する背景知識に基づき、その意味を特定する。また、上で挙げた例において、陸上の大会に関する新聞記事の中では、run が「人の高速移動」の意味として使用されやすいであろうという直感も、陸上大会に関する知識が、意味の特定を促進するためのドメインとして働いているからであろう。このように、動詞の意味と、語が用いられるトピックの関係は、動詞の意味とドメインの関係と（ほぼ）同じものとみなすことができるだろう。

上で述べた仮説は、一見すると自明のように思われる。実際、多義性の研究でトピックを暗黙裡に想定しているように解釈可能な研究もある。たとえば、Taylor (2003) は百科事典的意味論と分布意味論 (e.g., Turney & Pantel, 2010) の親和性を論じるに当たり、名詞 pen における多義性、すなわち筆記用具 (writing implements) と囲い (enclosures) の意味を例にとり、前者ならば直感的に ink, write, paper が共起しやすく、後者ならば children や動物の名前などが共起しやすいだろうと述べている。<sup>3</sup> Taylor の説明を上述の仮説の観点から言い換えると、筆記用具の意味は文房具に関するトピックで現れやすく、囲いの意味は家畜や懲罰などのトピックで現れる可能性が高いと考えることができる。

このように、Taylor (2003) などの先行研究はトピックの観点から解釈し直すことが可能である。しかし、ここで強調したいのは、これらの先行研究では明示的な形でトピックと多義性の関係を論じることはなかったという点である。すなわち、Taylor (2003) のような先行研究においては、語の多義性は主観的で曖昧に定義された (トピックとも解釈可能な) 文脈との関係に基づき説明されていたのである。

再現性が担保された形でトピックを論じるには、厳密なアルゴリズムに基づいてトピックを分類する必要がある。今日、コンピュータ技術の発展により、文書のトピックを確率的に計算する手法が考案されている。本稿では次節で紹介するトピックモデルと呼ばれる手法を援用することで上述の仮説の妥当性を検証及び論じていく。

### 3. リサーチデザイン

#### 3.1. トピックモデル

トピックモデルとは、ある文書がどのようなトピックなのかを解析する手法である。たとえば、次の例が何に関する文書かを考えてみよう。

- (2) 従来はペンと紙を持って椅子に座って考える研究者が多かった。しかし、近年はコンピュータの進歩、手法の発展により、膨大な資料の中から実際に使用される言葉の統計学的・確率論的な傾向を見出していくという研究スタイルが重要視されるようになった。

言語研究の歴史を知る者であれば、おそらく、(2) は (コーパス) 言語学に関する文書だと理解することができるだろう。しかし、本文にはトピックを表す「言語学」という語は含まれていない。(2) を (コーパス) 言語学に関する文書だと理解する者は、本文中の「言葉」や「研究」といった語から内容を判断していると考えられる。トピックモデルとは、語の分布や集合から文書がどのようなトピックと関係するものであるのかを統計的に解析し文書を分類する手法である。

近年のトピックモデルでは、Blei et al. (2003) が提案した latent Dirichlet allocation (LDA) を用いるのが一般的である (Müller & Guido, 2016, p. 348)。しかし、LDA は入力する文書データが短い場合には十分にトピックを学習することができず、結果としてトピック抽出の精度が下がるという問題が指摘されている (Tang et al., 2014)。3.3 節にて説明するが、本研究で使用する入力文書はストップワードを含めても 20 語にしかならないため、LDA での解析には不向きである。この問題を解消するために本研究では、短い入力データのトピック解析に適した biterm topic model (Yan et al., 2013, 以下, BTM) を利用する。

筆者らの知る限り、トピックモデルを語彙意味論研究に用いた先行研究はない。しかし、コンコーダンスラインをトピックモデルで解析することによって、ターゲット語がどのようなトピックで用いられているのかを調査することは理論的に可能である。本稿では、トピックモデルによって得られた結果をもとにコンコーダンスラインを精査することで、語のどの意味がどのトピックに現れるのかを調査することができることを示したい。なお、先に述べた BP 分析では、分析者が様々な ID タグを手作業によって付与していく必要があるが、本研究が用いるトピックモデルでは分析の部分はほぼ全て自動化されており、分析の客観性がより高い。

### 3.2. 使用するコーパス

本稿では 2 つのケーススタディを別々のコーパスを用いて行う。一つ目のケーススタディでは News on Web Corpus (Davies, 2016-, 以下, NOW コーパス) を使用する。本コーパスには 2010 年以降に公開された世界 20 カ国の英字オンラインニュース記事が収集されていて、2022 年 10 月 16 日の時点で約 160 億語が収録されている。すべてのデータを利用することは不可能なため、本研究ではフルテキスト版 2013 年米国データ (約 8,500 万語) に限定し調査を行う。

二つ目のケーススタディでは Corpus of Contemporary American English



(Davies, 2008-, 以下, COCA) を使用する。NOW コーパスと同様, COCA の全データを扱うことは不可能なため, ここでは, 1990 年から 2017 年までに出版されたフルテキスト版の Fiction データを利用する(約 1 億 1000 万語を収録)。ケーススタディ 1 とは異なるデータソースを用いることで, NOW コーパスとは別な角度からの分析が可能となることが期待できる。

### 3.3. 手順

本稿では, トピックと語の意味の関連を考察するにあたり, 以下の手順で調査を行った。まず, 共起語の取得範囲は, 前後 10 語, 合計で 20 語を抽出した。また, いくら BTM が短い文書解析に優れた手法とはいえ (3.1 節参照), 1 語や 2 語から構成される文書の解析は極めて困難なため, ストップワードを除外した後に 5 語未満となったデータを除外した。<sup>4</sup>

本研究ではデータ収集については, 従来のコーパスを用いた多義性の研究とは異なった収集方法を採用した。コーパスを用いた意味研究では, 文中に共起する語を調査するのが一般的である。たとえば動詞の意味研究の場合は主語や目的語, あるいはその修飾語などに注目することから, 共起語の検索は対象となる文を超えることはあまりない。しかし, 本研究は語が現れるトピックを扱うことから, 文を超えてデータ収集を行っている点に注意されたい。

以上の手続きに基づき, 本研究では英語の動詞 *run* をケーススタディとして扱う。本稿の冒頭でも述べたように, 英語の動詞 *run* は辞書に 30 以上もの意味が収録される語であり, これまでも百科事典的意味論の立場から様々な調査がなされてきた (e.g., Gries, 2006; Langacker, 1988)。本稿では, これらの先行研究との比較のために *run* を調査対象とした。

## 4. 分析結果

### 4.1. NOW コーパスにおける *run*

#### 4.1.1. 最適トピック数

トピックモデルでは, トピックの分類は自動的に行われるものの, トピックの数については分析者が決定する必要がある。最適なトピック数を決めるにあたり, LDA では, トピック数の妥当性を示す診断 (diagnosis) 等を利用する場合がある。しかし, BTM は文書の生成プロセスを計算しない手法であり, これらの指標を使用することができない (Yan et al., 2013, p. 1449)。ゆえに,



本研究では、排他性 (exclusivity) を利用することにした。

排他性とは、トピックモデルで計算したトピックの質を表す指標である。この指標は、トピックごとの単語の出現率 ( $\theta$  値) を、トピック全体の  $\theta$  値の和で割った値の平均値である。排他性は、値が高いほど、各トピックの特徴を表していると評価される。図1は、トピック数を5から50の範囲で5毎に計算し、各トピックにおいて、高出現率語10語を用いて排他性を算出した結果である。この図から、トピック数を15と40に設定したときに高い値を示していることがわかる。ここでは、これらのトピックを質的に解釈し、比較した結果、トピック数を15に設定するほうが解釈しやすい結果であったことから、排他性が最も高い値を採用した。

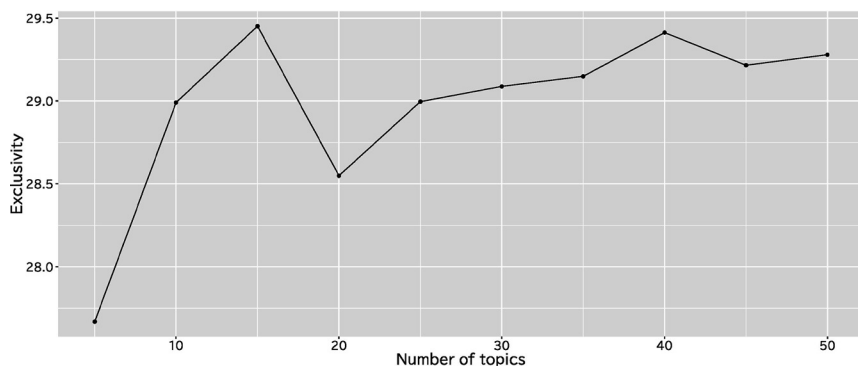


図1. Topic 5 から 50 の排他性 (上位 10 語を使用)

なお、BTMでは、計算の過程でテキスト中に幅広く使用される語を Topic 1 に集めることで、残りのトピックの精度が高められる。そのため、Topic 1 は研究においては利用されないトピックであり (background), 本稿でも Topic 1 は分析対象からは除外し、14 のトピックで調査を行った。

#### 4.1.2. BTM の結果

3.3 節で述べた手続きを通じて得たデータに対し、BTM で 14 のトピックを算出した結果を表1に示す (紙幅の都合により、一部抜粋)。この表の左端には各トピック番号を記載している。左から2列目には、排他性の算出時に用いた語彙を  $\theta$  値の降順に並べ、筆者らが解釈した各トピック名を3列目、そして各トピックで使用される run の意味については一番右の列に記述している。な

お、本研究では（フルテキスト版）コーパスに付与された POS タグを含めた調査を行っているが、本稿では視認性を高めるためそれらを除外して表示している。

表 1. トピックモデルによる分析結果

	トピック構成語上位 10 語...	トピック名	run の意味
Topic 2	power, gas, energy, fuel, natural, battery, plants, electricity, batteries, water	エネルギー	To execute/function
Topic 3	people, run, time, really, marathon, day, running, good, things, way	マラソン	Fast pedestrian motion
Topic 5	police, away, man, house, away, ran, toward, people, saw, old	犯罪	To escape
Topic 13	money, government, percent, fund, million, billion, social, dollars, security, deficit	政府と資金	To become used up
Topic 14	water, blood, run, face, hair, hot, hands, air, bar, room	液体の流れ	To flow
Topic 15	line, water, river, north, road, city, east, miles, aground, south	地理的概念	To flow (fictive motion)

次に、実際に各トピックにおいて run がどのように使用されているのかを確認していく。<sup>5</sup> 表 1 とはトピックの順番が前後するが、文字通りの移動 (literal motion) を表すトピックから見ていく。まず Topic 3 はマラソンに関するトピックだと思われる。Topic 3 で用いられる語と共起する事例を見てみると、マラソントピックで用いられる場合、run は「人の高速移動」である傾向が強い。

- (3) a. ... if you have not been exercising, don't expect to **run** a half marathon with only a couple weeks of running.
- b. The Boston Marathon is America's iconic race, the oldest marathon in the country, and the most important. Eighteen people **ran** it in 1897; last year, thirty-five thousand did.

次の Topic 5 は、犯罪に関わるトピックだとみなせる。このようなトピックでは、次のような「逃走 (to escape)」の意味で用いられる例が見られる。

- (4) a. ... there's the man whose picture you've seen countless times on TV over the past few days **running** toward your house as he returns fire to police officers in pursuit.

- b. Before Janelle turned 15, she was suspended from school, thrown out of her house by her parents, ran away from her foster home, and got arrested for shoplifting.

(4a) は、カーテンを開けた際に逃亡中の凶悪犯が警察に追われている状況を見たことが表されている。また (4b) では、Janelle が成長過程において様々な犯罪に手を染めてきたこと、そして彼女が里親から逃げた事象が述べられている。これらの事例で見られる用法は、先に見た「人の高速移動」と類似したものであり、文字通りの移動を表す点では同一の意味だと考えることができる。一方で、スポーツ、特に長距離走における移動では、誰から逃げるのか（追跡者）の存在が明示化されないが、(4) では警察や里親といった追跡者の存在が何らかの形で表される。このように (3) と (4) とでは、事態に関わる参加者が異なっていることがわかる。

次に Topic 2 では、natural gas や water power といった語が現れていることからエネルギーに関わるトピックであることがわかる。このトピックでの run は、Topic 3 や Topic 5 で見た文字通りの移動とは異なり、(5) の事例が示すように、「機械が作動する (to function)」という抽象的な意味で用いられている。

- (5) a. The devices run on natural gas and produce electricity with fewer emissions than a diesel gen-set.  
 b. Ordinarily water power would run the machines, but when the water wheel would not run on its own a steam engine would lift water from the stream …

Topic 14 と Topic 15 は、いずれも液体に関連するトピックであり、一見するといずれも「液体の移動 (movement of liquid)」の意味のように見えるが、実際の用例を見てみると興味深い差が見つかる。(6) は、Topic 14 の事例であるが、いずれの例も液体が移動する経路を表す前置詞 (from, down, over) が現れており、熱湯の移動を表している。このことから、Topic 14 の run は液体の移動事象を表していると言える。

- (6) a. Hot water from the pool above ran down the bank of the terrace, which was striated in several colors...  
 b. When it's time to serve, run hot tap water over the container to release the ice.

一方で Topic 15 の例に現れる語は、地理的な概念であり、これらが物理的に移動することは考えにくい。むしろ、概念化者が川や下水の経路を心的にたどった仮想移動 (fictive motion) (Talmy, 2000, cf., Langacker, 1986; Matsumoto, 1996) の一種だと考えるべきであろう。

- (7) a. Labeled “a state scenic river, a Catoctin Creek near Waterford runs under a single lane bridge on a lightly traveled road ...
- b. And yet its combined sewer system was constantly overflowing, in 400 places over 375 square miles, running afoul of federal water quality standards.

以上、トピックと語の意味の間には関係が成り立つことを示したが、ここで1つ疑問が生じる。本研究では14個のトピックを立てているため、14個のドメインを見出したことになるが、これらの背景知識は互いにどのような関係にあるのだろうか。本稿では、語の意味はトピックとの関係で決定されると考えることから (2.2 節参照)、ここではトピックの類似度を計算することで複数の意味間の関係を考えていく。図2は、確率分布の類似度を測る指標として Jensen-Shannon divergence を用いた結果を多次元尺度構成法 (MDS) に基づき描画したものである。

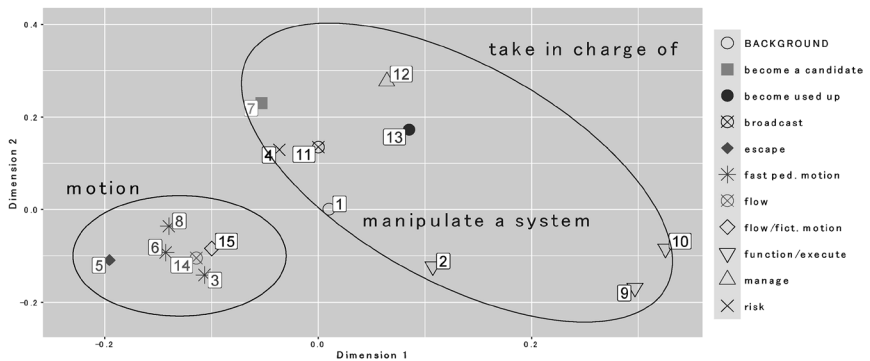


図2. MDSによるrunの意味の類似度 (1-2軸)

図2では、第3象限において「人の高速移動」(=Topic 3, 6, 8)や「逃走」(=Topic 5)といった文字通りの移動の意味が集まっている。また、Topic 14, 15の「物質の流れ」と「仮想移動」もそれらと同じまとまりに位置している。これらの

分布から、左下（青）の楕円で示す領域は run の「移動（motion）」の意味でまとめることができるだろう。ただし、2軸と3軸（図3）を見てみると、「物質の流れ」（Topic 14）と「仮想移動」（Topic 15）は第2象限に現れており、文字通りの移動の意味と離れたところに位置している。また、「逃走」（Topic 5）が「人の高速移動」（Topic 3, 6, 8）と離れて位置しているのも確認でき、各意味には一定の差があることが確認できる。図2の第4象限には、「機械が作動する（to function/execute）」の意味がまとまり、図2のY軸上の正の位置には「選挙に出馬する」や「経営する（to manage a company/business）」といった抽象的な意味が集まり、何かを「統括する（to take in charge of）」意味が集まっている。

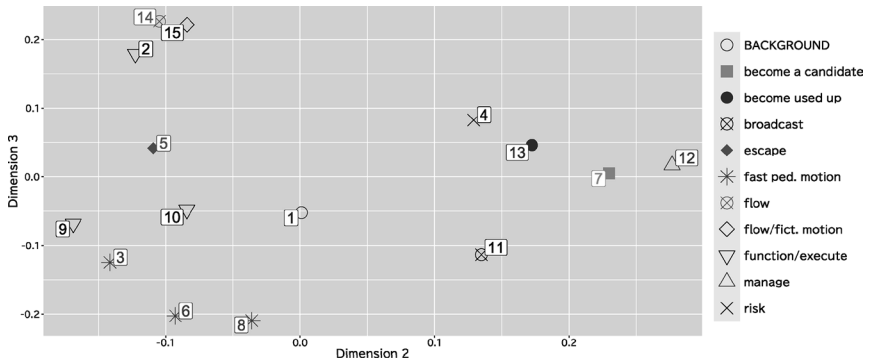


図3. MDSによるrunの意味の類似度（2-3軸）

以上の分布から、NOW コーパスをBTMで解析し、トピックからrunの意味を記述すると (i) 移動（motion）（図2の左下（青）の楕円）と (ii) 複雑なモノを機能させる（to cause complex entities to work）（図2の右（赤）の楕円）の意味に大別することができる。

## 4.2. COCAにおけるrun

### 4.2.1. 最適トピック数

Fiction データでは、4.1.1節と同じ条件で実験し、トピックの最適値を探った。図4は、その結果である。排他性は、トピック数が50の場合に最大値となっているが、数的に理想的なトピック数が必ずしも人間にとって解釈しやすいとは限らない（Chang et al., 2009; Törnberg & Törnberg, 2016）。そのため、本研

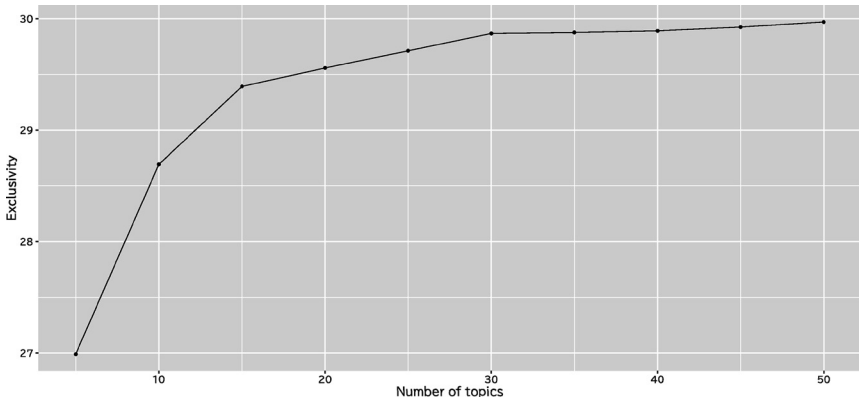


図 4. Topic 5 から 50 の排他性

究では、値の上昇が緩やかになる 30 から 50 の間を質的に解釈し、最もトピックの特徴が現れていたトピック数 40 を最適な数と判断した。

#### 4.2.2. BTM の結果

Fiction データを BTM で調査すると、4.1.2 節では見られなかった身体を表すトピックや建造物に関するトピックが確認できた。以下では、これら 2 つのトピックの特徴を見ていく。

まずは身体を表す表現について述べる。表 2 がそのリストである。Topic 11 の構成語を含む事例を見ると、(8) に示すように特定の感覚、あるいは感情へと変化の様子が描かれていることが分かる。そこで、この意味を「感覚・感情の瞬時的変化 (an instantaneous change of emotion/feeling)」と呼ぶことにする。寒気や震え、恐れといった感覚・感情は瞬時的に生じるものであり、この瞬時的な感情変化と run が持つ高速移動の様との類似性から、この表現が生じた可能性がある。

- (8) a. Erin grimaced in fear. I felt a tingle **run** up my spine.  
 b. He'd always envisioned her as [...] pale from fear or from the chill that **ran** along her spine.  
 c. She froze, her body **running** cold and clammy.  
 d. Zayder felt fear **run** in harsh prickles across his own skin.

表 2. Fiction における身体部位を含むトピック

	トピック構成語上位 10 語...	トピック名	run の意味
Topic 11	felt, spine, body, shiver, chill, moment, mind, cold, fear, eyes	感情や感覚	An instantaneous change of emotion/feeling
Topic 35	hand, finger, fingers, arm, left, scar, cheek, side, shoulder, neck	接触する身体部位と接触対象	To touch an object by moving a body part
Topic 37	hair, hand, fingers, hand, head, black, thick, brown, face, short	接触する身体部位と接触対象	To touch an object by moving a body part
Topic 39	face, tears, eyes, blood, hands, cheeks, sweat, mouth, nose, tongue	身体部位と液体	Body fluids flow on a body part

e. She closed her lips around my tongue, and a shiver ran through my body.

Topic 35 は、finger(s) や arm などの動かす対象の身体部位と、cheek や neck といった手で触れる身体部位を表す単語から構成される。そこで、本トピックを「接触する身体部位と接触対象の身体部位」とラベリングする。これらの語を含む事例には、身体部位の移動、特に (9a) のように「何かをなぞる」という意味を語用論的に示唆するものが含まれる。

- (9) a. I fetch my backpack, don't need unzip it cause the zipper don't work anyway, and pull out my book. It feel heavy and cool in my hand, and I run my finger cross the shiny silver letter of the title.
- b. She leaned in to rest her cheek against his and he ran a hand down her bare back, relaxing it above the concave of her spine.
- c. She palmed her neck. She ran her hands up and down her forearms, brushing up the gold hairs, then smoothing them again.

Topic 37 は fingers や hand といった身体部位を含む点で Topic 35 に類似している。しかし、接触対象となるものは、Topic 35 では (表 2 には現れていないが) letter (「文字」の意味) や scar といった身体部位以外の語彙も見られるが、Topic 37 は hair や brown, short といった髪の毛に関連する語や身体部位が多く含まれる。そのため、Topic 37 には「接触する身体部位と接触対象の身体部位」



というトピック名をつけた。これらの構成語が用いられる事例は次のようなものである。

- (10) a. She ran her fingers through her long, dark brown hair ...  
 b. ... Lou self-consciously ran a hand up and down the sides of his face.

本トピックと Topic 35 との間には、finger の文法的数の違いが確認できる。Topic 35 の場合、finger は単数形・複数形の両方が現れているが、Topic 37 では複数形しか現れていない。Topic 35 の場合、(9) で見たように、触れる対象が傷であったり文字であったりと、(もちろん、複数の指で触れるものもあるが)一般的に一本の指で触れるような接触対象物が含まれる。一方、Topic 37 の場合は髪の毛や頭など、指一本で触れることは稀であるものが含まれている。このように、run の意味を BTM で調査すると、対象物をどのように触れるのかという(いわゆる)言語外知識をも反映する結果を得ることができた。

表 2 の最後の Topic 39 は、hands や cheeks などの身体部位に加え、tears や blood, sweat といった体液が含まれている。これらの分布から、Topic 39 を「身体部位と体液」と判断できる。これらの語が使用される事例は次のとおりである。

- (11) a. Beads of sweat ringed his sunburnt head in a crown-like way and ran down into his watery faded-blue eyes.  
 b. You are crying, and your tears silently, uncontrollably run down your cheeks, along your nose, to your mouth.

(11) の事例が示すように、Topic 39 は人間の身体から出てくる液体についてのトピックであり、run は、「体液が身体部位を流れる」という意味で用いられていることが分かる。

以上のように、表 2 に挙げる全てのトピックにおいて、身体部位を表す表現が出ており、またトピック自体も身体に関わるものであることは間違いなさであろう。しかし、各トピックのコンコーダンスラインを観察すると、run の意味は、それぞれで異なっていることが明らかになった。特に、Topic 35 と 37 で確認した(いわゆる)言語外的知識の差に関しては、トピックの差異が語の意味の差を生んでいると言えるだろう。

次に、建造物に関するトピックについて見ていく（表3）。この表から、Fiction データにおける建造物に関するトピックでは、建造物（e.g., house, street）や内装（e.g., ext/int(erior), door）に関わる語が多く集まっていることがわかる。

表 3. Fiction における建造物に関わるトピック

	トピック構成語上位 10 語...	run の意味
Topic 3	toward, street, back, door, ext, night, man, starts, turns, people	Fast pedestrian motion
Topic 21	door, back, room, house, toward, front, came, stairs, kitchen, back	Fast pedestrian motion
Topic 24	int, night, room, ext, door, house, street, stairs, hallway, phone	Fast pedestrian motion

なお、これらの各トピックに固有のラベルをつけるのは非常に困難に思われる。なぜなら、構成語の分布が高度に類似しているからである。<sup>6</sup> そのため、本稿ではこれらのトピックについては「建造物」という総称的用語を用いて run の意味を調査する。建造物トピックでは、次のような run の事例が見られる。

- (12) a. I ran back through the dining room and toward the front door.  
 b. As Lisa lay there, shaken from the noise, she heard Aunt Mary call, “Lisa, I need your help!” Lisa ran down the stairs to the kitchen …

上記の事例が示すように、COCA の Fiction データにおいて、run が建造物トピックで用いられる場合、建造物が移動の始点あるいは着点となっている事例が目立つ。これらの例から、建造物トピックで現れる run の意味は「人の高速移動」であることがわかる。

この「人の高速移動」の意味は、NOW コーパスにおいても run の意味として確認することができたが（4.1.2 節）、コーパス間で言語表現として用いられる構文が異なっている。NOW コーパスでは、「人の高速移動」はスポーツの文脈で用いられることが多く、結果として（3）のように出場する競技や、run for 376 yards のように走る距離が後続する事例が多かった。一方、Fiction データだと、（12）に示したように run back や run through, run down A to B のように経路あるいは目的地が後続する事例が多く見つかる。この違いは、語の意味としては「人の高速移動」と分類可能であっても、トピックやレジスターが異なると、移動事象のどの部分に焦点を当てているのかが異なることを示している。<sup>7</sup>

## 5. 結果の解釈と結論

前節では、runの意味とトピックとの間に関連性があるという分析結果を報告した。より具体的には、BTMで得たNOWコーパスとCOCAのFictionデータとの分析結果を比較すると、共通して現れるrunの意味も存在するが、各コーパスの特徴を表すものも確認された。

まず、コーパス間で異なって見られた意味について考察する。COCAのFictionデータに特徴的な意味は、(9)や(10)に示した「身体部位を動かす」であった。これは、NOWコーパスを用いた本研究では確認されなかった用法であり、使用されるレジスターに偏りがあることを示唆している。Gries (2006)では、コーパスに含まれる様々なレジスターを統一的に扱っており、意味におけるレジスターの偏りについては言及していない。本研究により、レジスター間の意味の偏りを考慮に入れる必要性が示唆される。

次に、分析結果では、両コーパスに共通して現れる「人の高速移動」の意味においても重要な差が見られた。すでに論じたように、NOWコーパスでは目的語に出場競技をとる用法や、どのくらいの距離を走ったのかを示す例やトピックが目立ったのに対し、COCAのFictionデータでは経路を表す表現が目立った。これは「人の高速移動」という1つの事象に対して人が持つ百科事典的知識のうち、異なった側面を前景化していると言える (Fillmore, 1982 参照)。この点も、語の使用がレジスターに偏りがあることを示唆している。

以上の点は、2つのレジスターで言葉の使い方が明らかに異なっており、また使用される語の意味も異なっていることを示している。このことは、用法基盤モデルや百科事典的意味論においては、意味とレジスター間の深い関係性を論じる必要があることを強く示唆するものであり、言語の意味を社会・文化的観点から規定する必要性を主張する認知言語学の社会的転回 (social turn; Harder, 2010, p. 3) の方向性を支持するものである。

次に、方法論の点から、本研究の研究結果について述べておく。特に、BTMとGriesらが提案するBP分析との違いに着目して論じる。まず、BTMを用いることのメリットの1つに、分析に要する所要時間の短さと客観的信頼性の高さがある。BP分析では、例えば、動詞の意味を分析する場合には、相や時制に関する形態素タグ、動詞が生じる節や句の種類や文のタイプ (平叙文、命令文など) といった統語タグ、動詞の共起語の意味的特徴 (e.g., 主語や目的語が生物かどうか)、そしてターゲット語の意味などの情報が分析者によって

付与されていく。Gries (2006) では、815 の run の例に対して、252 個の ID タグが付与されているが、この内の 40% が手作業による形態、統語、意味のタグである。この作業は膨大な時間、労力がかかる上、タグの信頼性の確保が難しい。しかし、BTM であれば、これらの作業は不要である。これは、意味分類を手早く、また一定の客観的信頼性を保ちながら行うことを目指す分析者にとっては利点であろう（ただし、両者は競合する手法というわけではなく、補い合うものであると考えられる）。

2 つ目に、トピックというこれまでの多義性研究ではあまり考慮されてこなかった変数を意味分析の中心に位置付けることは、意味とトピックの関係を新たに探究する上で意義深いものと思われる。本研究が論じたように、意味とトピックには不可分な関係があり、曖昧な概念となりがちな文脈をトピックの観点から捉え直すことにより、新しい意味研究を切り拓くきっかけとなるかもしれない。

3 つ目に、コロケーション統計との比較から、BTM の利点を述べる。T 検定やカイ二乗検定量を用いたコロケーション分析では、例えば hair と fingers が run にとって重要なコロケーションだったとしても、run と hair、run と fingers のように 2 語の関係しか捉えることができず、3 語以上の関係を見ることはできない。しかし、BTM はトピックでの分析を行うため、hair と fingers が run とどのような背景知識を喚起するのかを分析することが可能である（4.2.2 節参照）。BTM が持つこの特徴は、言語に関わる世界の知識を統計的に浮かび上がらせる手法として有用だと言えよう。

なお、多くの研究と同様、BTM にも当然デメリットもある。まず、BTM や LDA は、実験のたびに結果が少しずつ変化するが、BP 分析は常に一定である（ただし、BP 分析でも、特に意味のタグ付けに関しては分析者が異なれば結果は一定とはならない）。また、BTM は、文書（あるいはコンコーダスライン）のトピック推定精度が低く、各トピックを構成する語を使って用例を探していく必要がある（注 5 参照）。一方、BP 分析では、すべてが手作業でタグ付与されるので、関連するデータを量的に調査することができる点については BP 分析に軍配が上がる。

本研究の結論は、以下のとおりである。表 1～表 3 に示したように、英語の動詞 run が使われるトピックを観察することができ、そしてそのトピックと動詞の意味との間に関連を見出すことができた。また、レジスター別のコーパスを利用し、動詞 run の用法を比較することで、語が前景化する百科事典的知識

の差も明らかにすることができた。BTMを用いることで、語の意味研究に加え、語の用法とレジスターの関係も分析することが可能となる。もちろん、BTMはまだ新しい技術であり、コンコーダンスラインの予測精度が低いという弱点はあるが、意味研究への新たな視点を導入することができる点で、有用な手法であると言えるだろう。

## 注

\* 本稿は、『英語コーパス研究』の3名の査読者より、有益なコメントをいただいた。記してお礼申し上げる。なお、本研究は、英語コーパス学会第47回大会での発表内容を発展させたものである。また、本研究は、科研費（基盤C研究課題20K00667）の助成を受けている。

- 1 本稿では、意義 (sense) ではなく、意味 (meaning) という用語を用いる。
- 2 すべての語の意味がトピックと関連するとは想定できず、前置詞などの文法化の進んだ語はトピックとの関連は薄いと思われる。この検証は今後の課題としたい。
- 3 Taylor (2003) の例は、同音異義語とも考えられるが、同音異義語と多義性との差は連続的であると考えられる (国広, 1982)。
- 4 トピックモデルでは、ストップワードを除外するべきではないとする立場もある (Brookes & McEnery, 2019)。しかし、ストップワードを含めた場合、筆者らが行った実験では、多くのトピックが機能語で構成され、解釈することができなかった。そのため、本稿では、ストップワードを除外した。
- 5 BTM では、トピック学習と文書トピックの推定は、別の手順で計算がなされる。具体的には、(i) コンコーダンスライン全体を入力データとしてトピックを学習し、(ii) (i) で学習したトピックを手がかりに、各コンコーダンスラインのトピックを推定する。表1や表2で示す結果は (i) の結果であるが、(ii) の推測精度が低い。そのため、(i) で得られた結果をもとに、分析者がコンコーダンスラインを探していく必要がある。
- 6 査読者より、各トピックに固有のラベルをつけるのが困難になった理由として、設定するトピック数が多すぎることが原因であると考えられ、「トピックの最適数を決める手法に問題は無いのか」という指摘を受けた。LDAでは、複数の指標を組み合わせて、総合的に判断するのが一般的であるが、BTMは、計算プロセスの特性から、最適なトピック数を決める数理的な手法は、あまり多く提案されていない (4.1.1節)。しかし、排他性は、BTMだけでなく、LDAなどでも広く使用されるため、一定の信頼性はあるものと考えられる。
- 7 コーパス言語学の伝統では、「テキストジャンル」と「レジスター」は、入れ替え可能な用語として用いられてきた。一方で、それぞれの用語に対して、一貫した定義が与えられてこなかった (Biber & Conrad, 2009; Conrad, 2015; Seoane & Biber, 2021)。そのため、本稿では、総称的な用語として「レジスター」を用いる。

## 参考文献

- Akita, K. (2012). Toward a frame-semantic definition of sound-symbolic words: A collocational analysis of Japanese mimetics. *Cognitive Linguistics*, 23(1), 67–90.
- Berez, A. L., & Gries, S. T. (2008). In defense of corpus-based methods: A behavioral profile analysis of polysemous *get* in English. *Proceedings of the 24th NWLC*, 157–166.
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Brookes, G., & McEnery, T. (2019). The utility of topic modelling for discourse studies: A critical evaluation. *Discourse Studies*, 21(3), 3–21.
- Bublitz, W., & Norrick, N. R. (2011). Introduction: The burgeoning field of pragmatics. In W. Bublitz & N. R. Norrick (Eds.), *Foundations of pragmatics* (pp. 1–20). Mouton de Gruyter.
- Cann, R. (1993). *Formal semantics: An introduction*. Cambridge: Cambridge University Press.
- Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Proceedings of the 22nd International Conference on Neural Information Processing Systems (NIPS'09)*, 288–296.
- Conrad, S. (2015). Register variation. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 309–329). Cambridge University Press.
- Croft, W. (1993). The role of domains in the interpretation of metaphors and metonymies. *Cognitive Linguistics*, 4(4), 335–370.
- Davies, M. (2008-). *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>.
- Davies, M. (2016-). *Corpus of News on the Web (NOW)*. Available online at <https://corpus.byu.edu/now/>.
- Divjak, D., & Gries, S. T. (2006). Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory*, 2(1), 23–60.
- Divjak, D., & Gries, S. (2008). Clusters in the mind? Converging evidence from near synonymy in Russian. *The Mental Lexicon*, 3(2), 188–213.
- Fillmore, C. J. (1982). Frame semantics. In The Linguistics Society of Korea (Ed.), *Linguistics in the morning calm* (pp. 111–137). Hanshin Publishing Company.
- Fillmore, C. J. (1985). Frames and the semantics of understanding. *Quaderni di semantica*, 6(2), 222–254.
- Gibbs, R. W., & Matlock, T. (2001). Psycholinguistic perspectives on polysemy. In H. Cuyckens & B. E. Zawada (Eds.), *Polysemy in cognitive linguistics: Selected papers from the international cognitive linguistics conference, Amsterdam, 1997* (pp. 213–239). John Benjamins.
- Glynn, D. (2009). Polysemy, syntax, and variation a usage-based method for cognitive semantics. In V. Evans & S. Pourcel (Eds.), *New directions in cognitive linguistics* (pp.

- 77–104). John Benjamins.
- Glynn, D. (2014a). The many uses of *run*: Corpus methods and socio-cognitive semantics. In D. Glynn & J. A. Robinson (Eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy* (pp. 117–144). John Benjamins.
- Glynn, D. (2014b). Polysemy and synonymy: Cognitive theory and corpus method. In D. Glynn & J. A. Robinson (Eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy* (pp. 7–38). John Benjamins.
- Glynn, D. (2016). Quantifying polysemy: Corpus methodology for prototype theory. *Folia Linguistica*, 50(2), 413–447.
- Glynn, D., & Robinson, J. A. (Eds.). (2014). *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*. John Benjamins.
- Gries, S. T. (2006). Corpus-based methods and cognitive semantics: The many senses of *to run*. In S. T. Gries & A. Stefanowitsch (Eds.), *Corpora in cognitive linguistics corpus-based approaches to syntax and lexis* (pp. 57–99). Mouton de Gruyter.
- Gries, S. T. (2010). Behavioral profiles: A fine-grained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon*, 5(3), 323–346.
- Gries, S. T. (2015). Polysemy. In E. Dąbrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (pp. 472–490). Mouton de Gruyter.
- Gries, S. T., & Divjak, D. (2009). Behavioral profiles: A corpus-based approach to cognitive semantic analysis. In V. Evans & S. Pourcel (Eds.), *New directions in cognitive linguistics* (pp. 57–75). John Benjamins.
- Gries, S. T., & Otani, N. (2010). Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal*, 34, 121–150.
- Haiman, J. (1980). Dictionaries and encyclopedias. *Lingua*, 50, 329–357.
- Harder, P. (2010). *Meaning in mind and society: A functional contribution to the social turn in cognitive linguistics*. Mouton de Gruyter.
- Heylen, K., Wielfaert, T., Speelman, D., & Geeraerts, D. (2015). Monitoring polysemy: Word space models as a tool for large-scale lexical semantic analysis. *Lingua*, 157, 153–172.
- Hoffmann, T., Horsch, J., & Brunner, T. (2019). The more data, the better: A usage-based account of the English comparative correlative construction. *Cognitive Linguistics*, 30(1), 1–36.
- Jansegers, M., & Gries, S. T. (2017). Towards a dynamic behavioral profile: A diachronic study of polysemous *Sentir* in Spanish. *Corpus Linguistics and Linguistic Theory*, 16(1), 145–187.
- Kishner, J. M., & Gibbs, R. W. (1996). How “just” gets its meanings: Polysemy and context in psychological semantics. *Language and Speech*, 39(1), 19–36.
- 国広哲弥. (1982). 『意味論の方法』. 大修館書店.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago Press.



- Langacker, R. W. (1986). Abstract motion. In *Berkeley linguistics society (BLS)* (Vol. 12, pp. 455–471).
- Langacker, R. W. (1988). A usage-based model. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics* (pp. 127–161). John Benjamins.
- Lederer, J. (2019). Lexico-grammatical alignment in metaphor construal. *Cognitive Linguistics*, 30(1), 165–203.
- Matsumoto, Y. (1996). Subjective motion and English and Japanese verbs. *Cognitive Linguistics*, 7(2), 183–226.
- Merriam-Webster. (n.d.). Run. Retrieved September 23, 2022, from <https://www.merriam-webster.com/dictionary/run>
- Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python*. O'Reilly.
- Nerlich, B., Todd, Z., Herman, V., & Clarke, D. D. (Eds.). (2003). *Polysemy: Flexible patterns of meaning in mind and language*. Mouton de Gruyter.
- Pustejovsky, J. (1995). *The generative lexicon*. The MIT Press.
- Ramsey, R. E. (2022). Individual differences in word senses. *Cognitive Linguistics*, 33(1), 65–93.
- Ravin, Y., & Leacock, C. (Eds.). (2002). *Polysemy: Theoretical and computational approaches*. Oxford University Press.
- Schönefeld, D. (2013). *It is ... quite common for theoretical predictions to go untested* (BNC\_CMH). A register-specific analysis of the English *go un-V-en* construction. *Journal of Pragmatics*, 52, 17–33.
- Seoane, E., & Biber, D. (2021). A corpus-based approach to register variation. In E. Seoane & D. Biber (Eds.), *Corpus-based approaches to register variation* (pp. 1–17). John Benjamins.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243.
- Stefanowitsch, A., & Gries, S. T. (2008). Channel and constructional meaning: A collostructional case study. In G. Kristiansen & R. Dirven (Eds.), *Cognitive sociolinguistics: Language variation, cultural models, social systems* (pp. 129–152). Mouton de Gruyter.
- Talmy, L. (2000). *Toward a cognitive semantics, volume 1: Concept structuring systems*. The MIT Press.
- Tang, J., Meng, Z., Nguyen, X., Mei, Q., & Zhang, M. (2014). Understanding the limiting factors of topic modeling via posterior contraction analysis. In *Proceedings of the 31st international conference on machine learning* (pp. 190–198).
- Taylor, J. R. (2002). *Cognitive grammar*. Oxford University Press.
- Taylor, J. R. (2003). Polysemy's paradoxes. *Language Science*, 25(6), 637–655.
- Taylor, J. R. (2012). Contextual salience, domains, and active zones. In H.-J. Schmid (Ed.), *Cognitive pragmatics* (pp. 151–174). Mouton de Gruyter.

- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins.
- Törnberg, A., & Törnberg, P. (2016). Combining CDA and topic modeling: Analyzing discursive connections between Islamophobia and anti-feminism on an online forum. *Discourse & Society*, 27(4), 1–22.
- Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37, 141–188.
- Tyler, A., & Evans, V. (2003). *The semantics of English prepositions: Spatial scenes, embodied meaning, and cognition*. Cambridge University Press.
- Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. In *Proceedings of the 22nd international conference on world wide web* (pp. 1445–1456).

(木山 直毅 北九州市立大学)

(渋谷 良方 金沢大学)