



コサイン類似度による英語の経路動詞の意味分析可能性を探る

森下, 裕三

(Citation)

統計数理研究所共同研究レポート, 465:92-104

(Issue Date)

2023-03-10

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/0100479389>

(URL)

<https://hdl.handle.net/20.500.14094/0100479389>



コサイン類似度による英語の経路動詞の意味分析可能性を探る

森下 裕三(桃山学院大学)

ymorishi@andrew.ac.jp

Exploring the Possibility of Cosine Similarity in Semantic Analysis of English Path Verbs

MORISHITA, Yuzo (St. Andrew's University)

Abstract

This study uses a distributional semantics approach to analyze English path verbs, with the goal of finding semantic differences between deictic and non-deictic motion verbs. The data set used was from the British National Corpus and consisted of 10 rows and 15,867 columns. The results, analyzed using cosine similarity, showed that deictic path verbs have distinct semantic properties compared to other path verbs.

キーワード

英語, 意味ベクトル, 経路動詞, 移動表現

1. はじめに

語がどんな意味をもつのかはどのようにして決まるのか。語の意味は「使用」によって決まるという考え方がある。この考え方は、Wittgenstein (1953) に端を発するが、哲学のみならず、コーパスを利用した言語研究とも相性がよい。コーパスにおける語の分布から語の意味を推測することができるからである。この考え方は、Wittgenstein による分析に着想を得た Firth (1957) の研究や、アメリカ構造主義言語学の潮流の中で Harris (1954) が発展させてきた分布意味論にみることができ、現在もなお発展を続けている。

本研究では、言語学において長い歴史をもつ分布意味論という考え方に基づき、英語における経路動詞と呼ばれる動詞の意味を分析する。経路動詞とは、以下の (1) や (2) の例にあるように、移動の経路という概念を語彙化した動詞のことである。

- (1) a. Harry **went** into the room.
b. Ron **came** out of the room.
- (2) a. Harry **entered** the room.
b. Ron **left** the room.

(1a) や (1b) の例にみられる *go* や *come* といった動詞には、直示的経路概念が語彙化されており、(2a) や (2b) の例にみられる *enter* や *left* などの動詞には、非直示的経路概念が語彙化されている。なお、移動の様態を語彙化した *walk* や *run* のような動詞は非常に多いものの、英語はあまり経路移動を多くはもたない言語である。

移動という事象は普遍的にみられるため、世界のどの言語にも移動事象を言語化する方法があると考えられる。世界の諸言語を調査すると、移動事象を言語化する手段は大きく 2 種類に分類でき、移動事象の中核をなす経路概念を主にどの要素によって語彙化するかによって、多くの言語は i) 衛星枠付け言語 (*satellite-framed language*) か ii) 動詞枠付け言語 (*verb-framed language*) のどちらかに分けることができる (Talmy 1991, cf. Talmy 1985)。なお、この場合の衛星とは前置詞や不変化詞などを指す。英語は、(3) の例にあるように、動詞ではなく *down* や *into* といった不変化詞や前置詞によって移動の経路を語彙化することが多いため衛星枠付け言語に分類される。

- (3) a. Luna walked **down** the street.
b. Severus ran **into** the door.

ここで、経路概念が動詞以外の要素によって語彙化されることの多い英語という言語において、なぜ本研究では経路動詞に注目するのかについて説明する。既に述べたように、英語の経路動詞はあまり多くない。直示的経路概念を語彙化したものに *come* と *go* があり、起点、通過点、そして着点といった非直示的経路概念を語彙化したものに *leave*, *cross*, *enter* などがある。これらを合わせても英語の経路動詞は 30 程度しかない (Levin 1993)。しかし、数が少ないからといって注目に値しないということではない。近年の移動表現についての研究では、直示性という経路概念をその他の経路概念と区別すべきかどうかについてさまざまな議論が展開されている (e.g., Choi and Bowerman 1991, Matsumoto et al. 2017)。

本研究では、上記のような移動表現についての近年の議論を背景に、分布意味論に基づく分析においても直示的経路概念を語彙化した経路動詞である *come* と *go* が、非直示的経路概念を語彙化した経路動詞と性質を異にするのかどうかを明らかにする。具体的な分析方法については 3 節で詳述するが、本研究では *British National Corpus (BNC)* を利用する。これは、Hilpert and Saavedra (2017) による研究を含め、理論言語学で広く活用されているのが BNC だからである。

本研究の構成はつぎのとおりである。まず 2 節では、分布意味論、および移動表現の類型論がどのように発展してきたのかについて概観する。また、英語の移動動詞の意味を分析する上で、分布意味論が役立つと考える根拠も述べる。つづく 3 節では、本研究で分析するデータおよび分析方法について詳述する。分布意味論といっても、さまざまな分析手法が存在し、語の意味に注目したものから、文書のトピックを明らかにしようとするものまである。したがって、どの分析手法を採用するかは、分析対象とあわせて検討する必要がある。4 節では、分析データの結果から理論的に

のようなことが言えるのかについて説明する。最後に 5 節では、本稿の内容を改めて振り返り、本研究をどのように発展させていけばより理論的な貢献につながるのかという今後の展望についても述べた上でまとめとする。

2. 先行研究

本節では、分布意味論がどのような言語観から生まれてきたのかを概観し、さらに、なぜ分布意味論という手法が移動動詞の意味分析にも応用できるのかという理由について述べる。

2.1 分布意味論の萌芽と発展

分布意味論の萌芽には、ふたつの異なる潮流がある。ひとつは、Wittgenstein (1953) による意味使用説に着想を得た Firth (1957) によるものである。この流れは現代のコーパス言語学にも引き継がれている (e.g., Sinclair 1991: 6)。もう一方は、アメリカ構造主義言語学のなかで Harris (1954) が示した分布意味論である。具体的な例として、彼は *oculist* という語と *eye-doctor* という語が似た環境に現れることを示し、現代のコーパス言語学における類義語研究にも通じる分析を示している (Harris 1954: 156)。本研究では、これらヨーロッパとアメリカにおいて発展してきた研究を区別せず、どちらも分布意味論として扱う。

分布意味論は、その後、語の分布をベクトルとして捉えた Jones (1986) による研究を土台として発展を続けている。また、近年は理論言語学の分野でも、語の多義性を考慮した分析手法が提案されている (e.g., Hilpert and Saavedra 2017, cf. Schütze 1998)。

本研究では、3 節で詳述する 2 種類の直示的経路動詞と 8 種類の経路動詞の語を分析対象とする。また、本研究では、計算量が大きくなることを避けるために、自己相互情報量 (Positive Pointwise Mutual Information) を計算するという方法を取る。

2.2 移動動詞の意味と分布意味論

移動表現の類型論は、移動事象を構成する中核的概念のうち、経路 (Path) という概念が動詞によって語彙化されるか、それとも動詞以外の付随的な要素によって語彙化されるかによって世界の諸言語を類型化しようというものである。現在では、Matsumoto et al. (2017) らによってさらに研究が進められており、移動表現の類型論も一枚岩ではない。本研究では、Matsumoto et al. (2017) らによる主張に基づき、経路概念を語彙化していると考えられる英語の動詞を上述の分布意味論の手法に基づいて分析する。

なぜ、経路概念を語彙化した動詞が重要なのかという点については上述の通りだが、Matsumoto et al. (2017) らは、とくに経路概念のうち直示的経路概念を他の経路概念と区別している (cf. Choi and Bowerman 1991)。直示性という概念を特別な経路概念だと考える理由はいくつかあるが、本研究では、つぎのような理由から直示的経路概念を他の経路概念と異なると考えている。

英語では、先にあげた (3) の例にあるように、不変化詞などの付随的要素によって経路概念が

語彙化されることが多い。しかし、実際には、(4) や (5) の例にあるように、動詞によって経路概念が語彙化されることも少なくない (Morishita 2014)。

- (4) a. When he **came** to the hospital after seeing Buster he was in floods of tears [...]. (BNC-CH1)
b. She switched off the CD player, **went** into the bedroom [...]. (BNC-A0R).
(5) a. He **arrived** in Canterbury on 27 May 669, [...]. (BNC-G0G)
b. When the baby was born, the maternity nurse had **departed** [...]. (BNC-A0R).

では、なぜ英語では経路概念を付随的要素だけでなく、動詞によって語彙化されることが珍しくないのだろうか。そして、先行研究で議論されてきたように、直示的経路概念が語彙化された動詞 (e.g., come, go) と非直示的な経路概念が語彙化された動詞 (e.g., arrive, depart) の意味が異なると考えるのであれば実証する必要がある。量的転回 (quantitative turn) を迎えた認知言語学において、質的研究によって主張されてきた理論的な仮説をコーパスなどの量的研究によって検証していくというのは重要な研究課題だと考えられる (e.g., Janda 2013)。

3. 残された課題の解決に向けて

これまで述べてきたように、英語という言語において、異なる 2 種類の経路概念を動詞が語彙化するという謎を解決する必要がある。言い換えると、直示的経路概念が、非直示的経路概念と異なる意味や機能を持つという仮説を実証する必要がある。

Harris (1954) や Firth (1957) が指摘してきたように、語の意味は周囲に現れる語によって明らかにできるという仮説が正しければ、分布意味論によって、英語における 2 種類の経路動詞の意味や機能の違いも明らかにできるはずである。以下では、このような考え方にに基づき、i) 分析する語、ii) 分析データ、iii) そして分析手法を紹介する。

3.1 分析対象とする語

これまで述べてきたように、本研究では、経路概念を語彙化した動詞の意味や機能の違いを明らかにすることが目的である。そのため、分析の対象となる語もすべて経路動詞と呼ばれるものである。しかし、経路動詞の数は多く、Levin (1993: 263–264) は 20 種類以上の動詞をあげている。しかし、これらの動詞の中には、つぎの (6) や (7) の例にある **plunge** や **recede** のように、物理的移動をあらわす自動詞の用例があまりみられないものも含まれる。

- (6) [...] she **plunged** into the water next to him. (BNC-JY4)
(7) The group of men **receded** and their voices faded. (BNC-C98)

本研究のような手法を利用する場合、コーパスにおいてある程度以上の頻度が得られる動詞を分

析する必要がある。したがって、本研究では、つぎの 10 種類の経路動詞のみを分析の対象としている。

- (8) a. come, go
b. arrive, cross, depart, enter, escape, leave, pass, reach

(8a) にあげた come と go は、直示的経路動詞で、(8b) にあげた 8 種類の動詞は Levin (1993: 263–264) が非直示的経路動詞としてあげているもののうち、分析に耐えうるだけの頻度が得られるものである。本研究では、これら 10 種類の動詞の意味および機能の違いを明らかにする。

3.2 データ

本研究で採用する分析手法は、これまで述べてきた分布仮説と呼ばれる考え方に基づいた意味分析である。ただし、分布仮説に基づいて任意の語や語彙を分析するための手法は非常に多く、研究目的によってどのような手法を取るべきかが大きく異なる。たとえば、Mitchell and Lapata (2010) による句や文の意味を分析する方法や、Turney and Pantel (2010) による文書のトピックを分析する方法などがある。一般的に、語のように小さな単位を分析する場合には、分析対象となる語の近くに生起する語の分布のみに注目し、より大きな文や文書のトピックを分析する場合には、分析対象となる語からやや離れた位置に生起する語の分布にまで注意を払う必要があると考えられている。

本研究で分析の対象としているのは経路概念を語彙化した動詞なので、上で述べた説明にしたがえば、任意の語の近くに現れる語の分布だけが重要だと考えられる。しかし、本当に語の意味を分析するのに前後 5 語程度の範囲で十分なのだろうか。詳細は 4 節で議論することになるが、直示的経路動詞と非直示的経路動詞の区別には、主語の位置に生起する名詞句の種類が重要になってくる。そして、主語の位置に生起する名詞句は、動詞から離れた位置にあらわれることも少なくない。たとえば、つぎの (9) にあげた例は BNC で見つけた come を含むものだが、動詞と主語の位置に生起する名詞句は 5 語以上離れている。

- (9) a. The lady prioress, a pure wool cloak wrapped around her, **came** down
to bid us adieu (BNC-HU0)
b. The school matron, followed by five teachers, three women and two men,
came rushing into the room. (BNC-CH4)

一般的な英語のコロケーションの分析では、任意の語から左右 4 語の位置に重要な要素があらわれるという意見もある (Stubbs 2002: 29)。しかし、本研究では、上述したように意味を考慮に入れた分析を目指すという理由から、語の前後 10 語以内に生起する語の分布を調査対象とした (cf. Begagić 2013)。ただし、代名詞 (e.g., he, she, I) や助動詞 (e.g., will, could), それに冠詞

(e.g., a, the) や疑問詞 (e.g., what, how) など各語の意味に大きな影響を与えないと考えられる語は排除した。しかし、後述するように、たとえば **Heathrow** という語は **come** や **go** よりも **arrive** や **leave** と共起する確率が高いなど、固有名詞の分布は語の意味の解明にとって重要な手がかりとなる可能性がある。こうした理由から、本研究では地名や国名だけでなく人名も分析に加えている。また、分布意味論では排除されることの多い前置詞も移動動詞の分析では重要な要素だと考え分析に加えている。その上で、つぎの表 1 のように各語と共起する語をベクトル化した。

表 1

英語の経路動詞と共起語の粗頻度

	Aaron	[...]	about	[...]	Heathrow	[...]	protected	[...]	Zurich
arrive	0		499		32		5		4
come	13		6,647		16		40		8
cross	0		190		0		4		1
depart	0		35		12		1		2
enter	0		384		0		17		1
escape	0		118		0		3		0
go	22		10,530		21		46		15
leave	9		2,312		28		28		10
pass	1		558		4		10		3
reach	0		771		6		8		2

表 1 は、既定の範囲内において 10 種類の経路動詞と 1 回以上の共起が確認できた語との共起頻度が記されている。このベクトルは 10 行 15,867 列におよぶ膨大なものだが、上述した一部の語を取り除く作業以外はすべて自動的に実施可能である。ただし、粗頻度だけでは各経路動詞がどれだけ意味的に類似しているのかわからないため、BNC 全体における各語の頻度に応じて、それぞれ語と語の組み合わせに重み付けをしていく必要がある。本研究では、正の自己相互情報量 (Positive Pointwise Mutual Information) を計算し、その計算結果に基づいて各動詞の類似度を計算した。以下では、具体的な分析手法について段階的に説明していく。

3.3 手法

本節では、前節で述べた手法によって収集したデータをどのように分析したのかについての詳細を述べていく。

3.3.1 自己相互情報量

前節までは、経路動詞と共起する語との粗頻度しか示していなかったが、より重要な共起関係を

つまびらかにするには粗頻度だけでは不十分である。たとえば、先に述べた **Heathrow** という語は、粗頻度だけを見ると **come** との共起が 16 なのに対して **depart** との共起は 12 である。しかし、**come** はこのベクトル全体では 2,109,265 の頻度があるのに対して、**depart** は 11,760 しかない。この数字だけをみてもわかるように、**Heathrow** は **come** よりも **depart** と共起しやすい。このように、より興味深い共起をみつけだす方法はいくつもあるが、本研究ではベクトルを利用した計算に基づく意味の推定において、もっともよく使われる方法のひとつである自己相互情報量を採用した。これは、つぎの (10) にある数式によって計算できる自己相互情報量 (Pointwise Mutual Information) の計算結果から負の数を 0 としたものである。

$$(10) \quad PMI(x, y) = \log_2 \frac{p(x, y)}{p(x)p(y)}$$

この数式で計算された値が、大きければ大きいほど、それらの語と語の共起関係はより意味のあるものだと考えられる。各動詞と共起する語との自己相互情報量を計算したものが表 2 である。

表 2

英語の経路動詞と共起語の自己相互情報量

	Aaron	[...]	about	[...]	Heathrow	[...]	protected	[...]	Zurich
arrive	0		0.07		3.76		0		0
come	0.33		0.03		0		0		0
cross	0		0		0		0		0
depart	0		0		6.07		0		0
enter	0		0		0		2.11		0
escape	0		0		0		0		0
go	0.60		0.20		0		0		0.53
leave	0		0		1.042		0.67		1.68
pass	0		0		0		0.67		0
reach	0		0		0		0		0

表 2 の結果をみても、自己相互情報量が 0 である **Heathrow** と **come** の共起よりも、自己相互情報量が 6.07 である **depart** との共起の方がより意味のあるものだということがわかる。

3.3.2 各経路動詞の類似度

つぎに、表 2 で示した自己相互情報量のベクトルデータから各動詞の類似度を計算する。数値は、ベクトルを利用した意味分析において利用されること多いコサイン類似度を使って計算した。ある語の意味と別の語の意味が完全に同じであればコサイン類似度は 1 となり、ある語の意味と別の語の意味の違いが大きければコサイン類似度は 0 に近い値をとる。この計算結果をまとめたものが表 3 である。

表 3

英語の経路動詞の類似度

	arrive	come	cross	depart	enter	escape	go	leave	pass	reach
arrive	1	0.04	0.15	0.14	0.21	0.16	0.01	0.14	0.15	0.20
come	0.04	1	0.02	0.01	0.05	0.02	0.12	0.11	0.04	0.07
cross	0.15	0.02	1	0.04	0.11	0.13	0.01	0.07	0.16	0.19
depart	0.14	0.01	0.04	1	0.07	0.08	0.00	0.04	0.04	0.05
enter	0.21	0.05	0.11	0.07	1	0.13	0.01	0.14	0.31	0.25
escape	0.16	0.02	0.13	0.08	0.13	1	0.01	0.09	0.11	0.11
go	0.01	0.12	0.01	0.00	0.01	0.01	1	0.06	0.01	0.02
leave	0.14	0.11	0.07	0.04	0.14	0.09	0.06	1	0.13	0.17
pass	0.15	0.04	0.16	0.04	0.31	0.11	0.01	0.13	1	0.24
reach	0.20	0.07	0.19	0.05	0.25	0.11	0.02	0.17	0.24	1

そして、このコサイン類似度を正規化したものが表 4 である。

表 4

各経路動詞の意味的距離

	arrive	come	cross	depart	enter	escape	go	leave	pass	reach
arrive	0	0.87	0.52	0.55	0.32	0.48	0.97	0.55	0.52	0.35
come	0.87	0	0.94	0.97	0.84	0.94	0.61	0.65	0.87	0.77
cross	0.52	0.94	0	0.87	0.65	0.58	0.97	0.77	0.48	0.39
depart	0.55	0.97	0.87	0	0.77	0.74	1	0.87	0.87	0.84
enter	0.32	0.84	0.65	0.77	0	0.58	0.97	0.55	0	0.19
escape	0.48	0.94	0.58	0.74	0.58	0	0.97	0.71	0.65	0.65
go	0.97	0.61	0.97	1	0.97	0.97	0	0.81	0.97	0.94
leave	0.55	0.65	0.77	0.87	0.55	0.71	0.81	0	0.58	0.45
pass	0.52	0.87	0.48	0.87	0	0.65	0.97	0.58	0	0.23
reach	0.35	0.77	0.39	0.84	0.19	0.65	0.94	0.45	0.23	0

これまででは自己相互情報量やコサイン類似度という指標によって、英語の経路動詞の類似度を示してきた。しかし、コサイン類似度は、各ベクトルの内積によって得られた数値であって、各ベクトルの距離をあらわす数値ではない。このような理由から、コサイン類似度を距離に変換した。コサイン類似度を距離に変換する方法として、本研究では、Levshina (2014: 330) にならい、表 3 におけ

る最大値である 0.31 で各コサイン類似度を割り、さらにその値から 1 を引いた数値を距離とした。このような手順を踏むことで、すべてのコサイン類似度は、0 から 1 の範囲の値をとることになる。語の意味が近ければ近いほど、ある語と別の語の距離は 0 に近づき、語の意味が遠ければ遠いほど、ある語と別の語の距離は 1 へと近づく。

表 4 のデータは類似度を距離に変換したものなので、数値が大きいほど、それぞれの語の意味は異なる。したがって、たとえば go は arrive, cross, enter, escape, そして pass といった多くの経路動詞と意味が大きく異なるということになる。一方で、arrive は enter, reach, escape, pass, そして leave などの経路動詞と意味が近いことが明らかになった。

4. 結果と考察

本節では、分布意味論に基づく各経路動詞のコサイン類似度による意味の推定によって明らかになった点について考察する。そして、最後に具体的な BNC の用例を分析しつつ、本研究の結果が示す理論的意義についても議論する。

4.1 データの全体像

表 4 で示した各経路動詞の距離を示したデータによると、英語における経路動詞のうち、もっとも意味に近いのは距離が 0 となった enter と pass, 次いで距離が 0.19 となった enter と reach, さらに距離が 0.32 である enter と arrive とつづく。これらの動詞のうち、enter と reach, そして enter と arrive については、どれも着点という経路概念を語彙化しているため高い類似度を示すことは理解できる。しかし、pass と enter の類似度が高い理由は説明できない。なぜなら、pass は通過点という経路概念を語彙化しており、enter は着点という経路概念を語彙化しているからである。

このような結果になった原因を探るために、自己相互情報量のデータに立ち返る必要がある。表 2 を見ても、他の動詞とは違い、これらの動詞は Heathrow や Zurich といった移動動詞とは相性の良さそうな場所をあらわす名詞や、about といった前置詞との共起は重要でないようにみえる。一方、むしろあまり移動事象とはかかわりのなさそうな protected といった語との共起に強い結びつきがあるように思える。

これには、語の多義性が関与していると考えられる。つまり、pass という語は移動以外の意味で使われる傾向にあるということである。なお、移動以外の意味で使われる pass を含む例としては以下の (11) のようなものがある。

- (11) a. Also in December 1989 a revision of the Family Law **was passed** after 30 years
Of campaigning by women's groups. (BNC-HKR)
- b. The river **passed** at its feet, pulling at the reeds along the bank. (BNC-FEM)

(11a) の例にみられる pass は、議案や法案などが通過するという意味であって物理的な移動を

あらわしているわけではない。また、(11b) の例にみられる *pass* は、虚構移動と呼ばれるもので物理的に川が移動しているということを意味しない (Talmy 1996, Matsumoto 1996)。

そして、もっとも興味深いのが、本研究の主題ともかかわる直示的経路動詞のうち、*go* がすべての非直示的経路動詞ときわめて異なる性質を示しているということである。表 4 のデータを見ると、*go* は *come* との距離が相対的に近く 0.61 を示すものの、その他の動詞との距離は 0.81 から 1 とかなり離れていることがわかる。非直示的経路動詞のなかでもっとも *go* との距離の近い *leave* でさえ距離は 0.81 にすぎず、かなり意味の異なる動詞ということになる。

4.2 直示的経路動詞の特殊性

ここで、意味的な距離が近いか遠いかの境界を 0.5 とし、本研究で問題にしている直示的経路動詞である *come* と *go* が他の経路動詞とどのような関係にあるのかをみることにする。まず、特徴的なのは *go* である。この語は、意味的な距離が近い語がまったく存在せず、先に述べたように、もっとも近い *come* ですら距離は 0.61 である。そして、ほぼ同じことは *come* についても当てはまる。なぜなら、*come* も *go* と同じく意味的な距離が近い語が存在しないからである。なお、他の経路動詞をみてみると、*depart* も意味的な距離が近い語を持たない語でありやや特殊な経路動詞ということになりそうである。これらの事実から、分布仮説に基づくベクトルを利用した意味分析においても、やはり *come* と *go* は他の経路動詞とは性質が異なるということが明らかになった。

では、なぜ周辺に生起する語の分布という情報だけで、直示的経路動詞を非直示的経路動詞から区別することができたのだろうか。ここで、ふたたび *come* と *go* の自己相互情報量を確認してみることにする。既に述べたように、自己相互情報量というのは、ある語と別の語との共起関係が特別な意味をもつと考えられるようなときに高くなる値である。本研究でいえば、他の経路動詞とはあまり共起しないが、*come* や *go* といった直示的経路動詞と共起しやすい語がどのようなものかが自己相互情報量のデータからわかるということである。このような考え方に基づいて、*come* と *go* の周囲に分布する特徴的な語を調べてみると、両者に共通して自己相互情報量が比較的高い語が見つかった。それらの語を含む例を (12) にあげる。

- (12)a. Well wait for Aaron to **come** out and you, well you go in [...]. (BNC-KD1)
b. Two weeks later Jason **went** to Scotland's biggest-ever rave. (BNC-ACN)

これらの例をみてもわかるように、直示的経路動詞は非直示的経路動詞よりも固有名詞と共起しやすいという傾向がある。このことは、自己相互情報量を示した表 2 において、固有名詞の *Aaron* が *come* と *go* において高い値を示していることからわかる。

では、なぜ *come* と *go* は他の非直示的経路動詞よりも固有名詞と共起しやすいのだろうか。それは、i) これらの語が他の経路動詞とは異なり自動詞としての用法しかもたない語であること、さらに、これは *come* のみに言えることではあるが、ii) 心理的近接性が感じられる場合に使われやすいということが関与している。以下では、これらについて順に説明していく。

まず、自動詞か他動詞かという問題について議論する前に、Du Bois (2002) による選好的項構造 (preferred argument structure) と Ariel (2008) による指示表現の階層について簡単に説明しておきたい。コーパスに生起する文というのは、一部の例外を除いて、書きことばであれ話しことばであれ、先行する文もあれば後続する文もある。言い換えると、基本的に文は談話のなかの一部ということになる。このような談話において、新情報は自動詞の主語や他動詞の目的語として生起しやすく、旧情報は他動詞の主語として生起しやすいという傾向がある。これが選好的項構造という考え方である。この考え方を発展させた Ariel (2008) によれば、不定冠詞をともなうような不定記述 (indefinite description) や固有名詞などは新情報、あるいは不活性的とされ、定冠詞をともなう定記述 (definite description) や代名詞は旧情報、あるいは活性的とされる。つまり、談話の構造上、英語の自動詞の主語の位置には、名詞句は不定記述や固有名詞が生起しやすいということである。

ここまでの説明でわかるように、直示的経路動詞が固有名詞と共起しやすいというのは、選好的項構造や指示表現の階層構造という理論的前提に裏付けられた性質だと考えることができる。pass や cross などの非直示的経路動詞は、他動詞としての用法をもつため、主語の位置にあまり新情報である固有名詞は生起しない。これが、周囲に現れる語の分布が、直示的経路動詞と非直示的経路動詞とで異なる理由のひとつである。

つづいて、心理的近接性と come との関係について議論したい。Matsumoto et al. (2017) は、come が直示的性質のみならず、心理的近接性をあらわす時に使われやすい傾向を示すと述べている。文のなかで固有名詞が使われるのは、その人物について読み手が知っていることが前提となる。たとえば、読み手がまったく知らない人物に言及する場合、一般的には人名などの固有名詞で言及するのではなく、不定冠詞をともなう名詞句などが使われる。このように考えると、come が固有名詞と共起しやすいというのは、妥当な分析だと言えそうである。

5. まとめと今後の課題

本研究は、任意の語の周囲に生起する語の分布情報のみから、語の意味の違いを明らかにすることができるかどうかを検証することを目的としたものである。先行研究で議論されてきた、経路動詞における直示的性質と非直示的性質との区別という意味や機能の違いを自動的に分析できるといえるのは、意味分析の可能性を広げるものだと言える。

意味を分析するためには、人間が目視でひとつひとつの用例をつぶさに観察することでしか実現できないと考える研究者もいる (Divjak and Fieller 2014: 208)。しかし、本研究では、自動的意味分析もある程度まで可能であることを示唆する結果を得た。ただ、今回の分析結果が必ずしも最適な分析だとは言えない。なぜなら、4.2 節で説明した選好的項構造や指示表現の階層構造が、直示的経路動詞と非直示的経路動詞の意味の区別にとって重要なのであれば、冠詞や代名詞といった一般的にベクトルを利用した意味分析で排除される要素を残しておくべきだった可能性があるからである。しかし、今後の研究でより優れた分析モデルを構築するための土台として本研究における発見は意義をもつ。固有名詞との共起が直示的経路動詞と非直示的経路動詞の区別に

かかわるというのは、これだけ膨大なデータを分析することによって初めて見えてくるものだからである。

言語研究における多くの理論的前提を十分に考慮に入れた上で、大量のデータを分析し、分析のためのモデルの妥当性を量的にも質的にも検討し、さらに、検討した結果をより優れたモデルの構築のために活用する。こうしたサイクルをつづけながら、言語研究における理論的な貢献へと繋げることが今後の課題である。

謝辞

本研究は科学研究費プロジェクト「英語の移動表現における経路概念の機能的分析」(20K13069) による研究成果の一部である。

引用文献

- Ariel, M. (2008). *Pragmatics and grammar*. Cambridge University Press.
- Begagić, M. (2013). Semantic preference and semantic prosody of the collocation make sense. *Jezikoslovlje*, 14(2), 403–416.
- Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns, *Cognition*, 41, 83–121.
- Du Bois, J. W. (2002). Discourse and grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure*, vol.2 (pp. 47–87). Lawrence Erlbaum Associates.
- Divjak, D., & Fieller, J. N. (2014). Cluster analysis: Finding structure in linguistic data. In D. Glynn & J. Robinson (Eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy* (pp.405–441). John Benjamins Publishing Company.
- Firth, J. R. (1957). A synopsis of linguistic theory 1930–1955. *Studies in Linguistic Analysis*, 1–32.
- Harris, Z. (1954). Distributional structure, *Word*, 10, 146–162.
- Hilpert, M. & Saavedra, D. C. (2017). Using token-based semantic vector spaces for corpus-linguistic analyses: From practical applications to tests of theoretical claims. *Corpus Linguistics and Linguistic Theory*, 16, 1–32.
- Janda, L. A. (2013). *Cognitive linguistics: The quantitative turn: The essential reader*. De Gruyter Mouton.
- Jones, K., S. (1986). *Synonymy and semantic classification*. Edinburgh University Press.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. The University of Chicago Press.

- Levshina, N. (2014). *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins Publishing Company.
- Matsumoto, Y. (1996). Subjective motion and English and Japanese verbs. *Cognitive Linguistics*, 7(2), 124–156.
- Matsumoto, Y., Akita, K., & Takahashi, K. (2017). The functional nature of deictic verbs and the coding patterns of deixis. In I. Ibarretxe-Antuñano (Ed.), *Motion and spaces across Language: Theory and applications* (pp.95–122). John Benjamins Publishing Company.
- Mitchell, J., & Lapata, M. (2010). Vector-based models of semantic composition. *Proceedings of ACL-08*, 236–244.
- Morishita, Y. (2014). A quantitative constructional approach to converbial motion constructions in English. [Unpublished doctoral dissertation]. Kobe University.
- Schütze, H. (1998). Automatic word sense discrimination. *Computational Linguistics*, 24, 97–124.
- Stubbs, M. (2002) *Words and phrases: Corpus studies of lexical semantics*. Blackwell Publishers.
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description*, vol.2 (pp.36–149). Cambridge University Press.
- Talmy, L. (1991). Path to realization: A typology of event conflation. *Proceedings of the seventeenth annual meeting of Berkeley Linguistic Society*, 17, 480–519.
- Talmy, L. (1996). Fictive motion in language and ‘conception.’ In P. Bloom, M. A. Peterson, L. Nadel & M. F. Garrett (Eds.), *Language and space* (pp.211–276). MIT Press.
- Turney, P. D. & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37, 141–188.
- Wittgenstein, L. (1953). *Philosophical investigations*. Blackwell Publishers.

Corpus

British National Corpus. Available from <http://www.natcorp.ox.ac.uk/>