

# コーパス分析によって明らかにする 新語によるコミュニケーションを可能にする複合名詞の語形成

大友和幸<sup>1</sup> 橋本敬<sup>1</sup>

<sup>1</sup>北陸先端科学技術大学院大学 知識科学系  
{s2110033, hash}@jaist.ac.jp

## 概要

言語の共通性と新規性が両立するメカニズムを解明するため、商標登録された新規な複合語を対象とし形式・意味の両面の語形成について分析した。形式における共通性・新規性を複合語とその構成語の出現数の推移より調査し、既存の構成語を結合させて新規な複合語が作られることを確認した。意味の共通性の分析として、対象語の共起語分布のジニ係数でその語がどれだけ限定的な（共通の）意味で使用されるかの定量化を試みた。意味の新規性として、個々の構成語にはないが、語が結合した複合語でのみ出現する創発的意味を調べる方法を2つ提案する。一つは共起語の出現回数の違いを調査する方法で、もう一つはLDAを使用して構成語と複合語に分配されるトピック群を比較する方法である。

## 1 背景

### 1.1 言語の共通性と新規性

コミュニケーションに用いられる言語は、共通性と新規性という2つの性質を持つ。[1, p233]

共通性とは、語の意味がおおむね決まったものとして使用され、それが使用者に共有されている、すなわち使用者により千差万別の意味を持たない性質を指す。信号理論[2]に基づくコミュニケーションの見方では、話し手が言語化（外在化）したメッセージを聞き手が理解するためには、コミュニケーションを行う双方が語彙や文法に関する知識を共有していることが必要となる。したがって、共通性はコミュニケーションの前提とされているといえるだろう。

その一方で、我々は、今までになかった新しい形式を用いて新しい概念を言語で表現することもできる。すなわち、言語は新規性を生み出す創造的な性質を持つ。ファッションコミュニティを対象にした

コーパス分析[3]では、ユーモア、皮肉、挑発に関連した意味に加え、説明、評価、増幅、緩和等を目的として、新たな概念をうまく表現するために新しい形式の語が出現していることが示されている。このように、既存の習慣化された語ではなく、今までに見ることがなかった新しい語の形式が誕生するということが実際の現象として確認されている。

共通性が習慣化したパターンによって曖昧さを下げることができるだけ正確な伝達を可能にする傾向であるのに対し、生成された新語は厳密には話し手しかその意味を知りえないため、新規性は正確な伝達を妨げる傾向にある。使用される語や文法に関する知識が双方に共有されていない場合には、自分の意図を正確に相手に伝えることができず、発話に表現される意図を伝達するという意味でのコミュニケーションを成り立たせることが困難となる。このように、共通性と新規性には相反する傾向性があるため、両方の性質を担保したまま新語の生成とコミュニケーションの両方を説明することができていない。

### 1.2 複合語が持つ意味の創発性

新語を生むもっとも良く使われる方法のひとつが、既存の語を組み合わせる新しい形式を作る複合語である[4]。複合語の意味は部分的合成性を持つと言われる[5, p.8]。すなわち、複合語の意味は構成要素の意味と無関係ではないが、単純に足し合わせたものでもないため、個々の構成語の単純な総和として複合語全体の意味を説明することができない。メンタルスペース理論における概念融合[6]はその典型である。たとえば「butcher surgeon」という表現は「下手な外科医」という意味を持ち得るが[7]、butcher「屠殺業者」が「下手な」という意味を持つわけではなく、surgeon「外科医」と組み合わせさせたときに生じる創発性を持つと考えられる。創発とは、全体が要素の集合以上の性質を持つことであり、ここでは複

合語の意味が構成要素の意味の単純な和ではないことである。

複合語が持つ意味の部分的合成性あるいは創発性から考えると、共通性と新規性が両立しつつコミュニケーション（伝達）を成立させるメカニズムを説明する仮説として、以下のものが考えられる。

習慣的に使用され共通性を持つ要素同士を組み合わせ、その組み合わせパターンによって複合語全体で新規性を持たせる。

すなわち、新しい概念を表現する際に、全く新しい語をあてがうのではなく、既存の語を組み合わせることにより新しい概念を表現する。このような語を使用することで、今までに聞いたことがない新語に対しても意味を理解し、コミュニケーションを成立させられる。

### 1.3 目的

本研究は共通性と新規性が両立するメカニズムの解明のため、上記の仮説の検証を行うことを目的とする。具体的には、複合語として多くの言語でよく用いられる複合名詞に限定し、新規に出現した複合名詞の形式と意味の両方において、共通性を持った語を構成語として結合することによって新規性を持つ複合語が作られていることを示す。

それとともに、意味面の分析において、意味的な新規性を評価し、既存の語が結合され複合語になる際に生じる複合語全体の意味の創発性を評価する方法を提案する。

## 2 手法

上記の目的を達成するため、新規に出現した複合語や、それを分解した構成語について、出現の推移を計測し、形式面での共通性・新規性を評価する。意味面の共通性として使われ方の偏りを共起語分布のジニ係数で評価する。意味の創発性を評価するために、複合語にはあるが各構成語にはない共起語を分析し、複合語のみに存在する意味の検出を試みる。また、LDAを使用した意味の検出手法を提案し、構成語が結合することによって付与した意味についての分析を行う。

### 2.1 複合語の取得

本研究の複合語は、なにかしらの意図伝達を目的とした新語であることが求められる。そのためには、対象独自の個別的な属性を明示しようとする意識で

ある「表現性」と、その所属先カテゴリを明示しようとする意識である「表示性」[8]が反映されている必要がある。これを満たすものとして、新しい商品の商品名が考えられる。商品名はパッケージなどに記載され、消費者がその商品を理解したうえで購入できるような属性などを示す必要があるため、表示性を持つ。また、商品名は消費者の注意を引くような名前自体の独自性や、その物自体の独自性を伝えることを目的として、他の商品とは違う名前にする必要があると考えられる。

本研究では、商標データベース（工業所有権情報・研修館が提供する特許情報プラットフォーム <https://www.j-platpat.inpit.go.jp>）にて「商標」を対象として検索）に登録されている消費者向けの商品名の中から複合語の商品名を取得し、それらの語の語形成について分析を行う。

### 2.2 形式面の分析方法

形式の面における共通性・新規性として以下の性質を評価する。

**共通性** 複合語の出現前から構成語が出現している

**新規性** 複合語はその商品が出現する前までは使用されていない

これらの性質を評価するためには、取得した複合語とその構成語それぞれについて出現の程度とその推移を調べる必要がある。すなわち、語の使用時期が特定でき、長期間に渡って観察できる分析データが必要である。また、今回扱う商品名はニュース記事や公文書等に出にくい、商品名が使用される場面のデータである事が求められる。

今回の研究では、語の出現の計測に Google Trend 人気度を使用した。ある語の Google Trend 人気度は、その語の検索数に応じた相対的な値を [0,100] の範囲で表示する。すなわち、その語について十分な検索クエリがない場合には 0、指定した期間内における最高値は必ず 100 となるよう算出される。Google Trend 人気度は月ごとに算出される。また、Google 検索は一般的に商品の情報取得や通信販売などで利用されている点から、商品名の出現の推移を計測するのに適していると考えられる。なお、算出対象とする全期間で十分な検索が無い場合はエラーとなるが、本稿では全期間の人気度を 0 とした。

## 2.3 意味面の分析方法

意味の分析においては、意味に共通性がある構成語を結合させることにより、複合語全体では各構成語にはない意味が出現することを示したい。ここでは、ある語の意味はその語がどのような使用のされ方に反映されると考える。以上の考えより、ある語の意味の面における共通性・新規性を以下の性質で評価する。

**共通性** 語がある程度決まった形で使用されており、使用者によって大きな違いがない

**新規性** 構成語が結合し複合語となるときに、個々の構成語にはない使われ方が出現する

そして、語の使用のされ方を共起語の分布で表す。

意味の共通性は、使用されている語の分布の偏りにより評価する。ある語が共通性を持ち限定的な意味で使用されている場合、特定の語と共に使用されると考えられる。その一方で、ある語が共通性を持たずに使用者や使用場面により様々な意味で使用される場合、使われる意味の数に応じて共起語が増え、十分な頻度がある共起語の数が多くなり、共起語分布の偏りは小さくなると考えられる。この分布の偏りはジニ係数で評価できるだろう。すなわち、ある語の出現回数でローレンツ曲線を計算すると、共通性が高いほど一様分布を表す完全平等線から乖離する。反対に、共通性が低いほど完全平等線に近くなると考えられる。

以上を踏まえ、本研究では、複合語・構成語それぞれのツイートにおける共起語を収集し、共起語の出現回数におけるジニ係数を計測することで、共通性を評価する。

意味の新規性として、構成語には無いが複合語全体にはある意味、すなわち創発的意味の検出を試みる。そのための方法として、共起語を用いる方法と潜在的ディリクレ配分法 (LDA) を用いる方法を提案する。

共起語を用いる方法として、複合語・構成語について、それぞれを含むツイートを取得し、最も多く共起した語を100語取得する。その100語について、構成語には出現しないが、複合語では出現する語を取得する。

LDAを用いる方法では、複合語・構成語が分配されるトピックを計測し、複合語では分配されるが構成語には分配されないトピックを観察する。トピックを構成する学習には取得した商品名と同じジャン

ルに属する Amazon の商品レビューを用いる。本稿では100トピックに分割した。

## 3 結果

### 3.1 形式面の分析結果

商標データベースに記載されている菓子で複合語になっている商品名について GoogleTrend 人気度の推移を分析し、仮説を裏付けると考えられる例を確認した(図1上「チョコまみれ」の例)。すなわち、構成語の「チョコ」「まみれ」の人気度は「チョコまみれ」の発売前(2020年5月)から人気度があり、複合語「チョコまみれ」は発売後のみで人気度が高まっている。だが取得した複合語商品名の多くは、構成語は人気度の推移が確認できる一方で、複合語の人気度は全ての期間で0となっていた(図1下「おいしいお菓子の御神木」の例)。

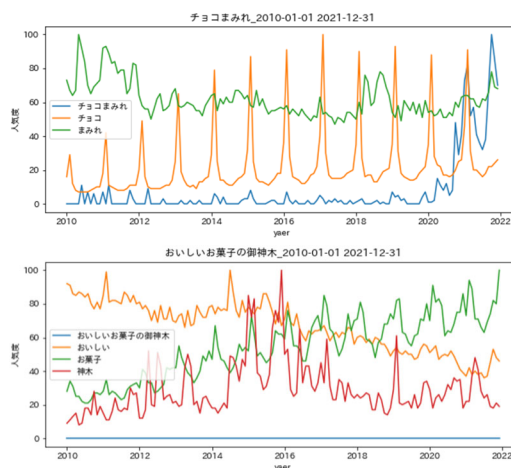


図1 構成語の共通性・複合語の新規性

### 3.2 意味面の分析結果

意味分析のため2022年12月31日以前の直近ツイートを5,000件取得した。Botや繰り返しのツイートによるデータの偏りを避けるため、全く同じ文章のツイートは1ツイートのみ分析に使用した。

ジニ係数による意味の共通性の分析を示す。分析対象は複合語「チョコまみれ」「美味しいお菓子の御神木」とその構成語、およびターゲット語の上位カテゴリである「菓子」を含むツイートの共起語の出現回数(共起語分布)のジニ係数を算出した(表1)。上位カテゴリの語は様々な文脈で使われるため、共起語分布の偏りは相対的に少なくジニ係数は低くなると考えたが、そのような結果にはならなか

った。意味の共通性に関して、複合語、構成語、上位カテゴリ語の共起語分布のジニ係数で比較してもあまり大きな差はない事がうかがえる。

表 1 分析対象語の共起語分布のジニ係数

単語	ジニ係数
菓子	0.75
チョコまみれ	0.65
チョコ	0.72
まみれ	0.68
おいしいお菓子の御神木	該当ツイートなし
おいしい	0.72
お菓子	0.75
神木	0.76

意味の新規性として「チョコまみれ」とその構成語の共起語を分析した結果、構成語に出現せず複合語では出現する語として表 2 の 21 語が確認された。その中には「だらけ」「美味しい」「初めて」など、一般的な語があり、「チョコまみれ」の上位カテゴリである「カントリーマアム」「クッキー」、発売元の「不二家」だけではないことが確認できた。LDA による意味の分析のため、Amazon で「菓子・スナック」カテゴリに属する商品の全レビューを 100 トピックに分割してトピックを学習した。分析対象語を 1 文書とみなし、その文書が分配されるトピックを観察した。「チョコまみれ」と「チョコ」の配分カテゴリは一致し、「チョコまみれ」のみに現れる意味の検出はできなかった

表 2 複合語とだけ共起する語

	該当語
チョコまみれ	だらけ, カントリーマアム, ホーム, 口, 袋, 最近, 美味しい, クッキー, 枚, うまい, 昨日, カントリーマアムチョコ, 感, 初めて, 朝, 限定, 缶, 明日, 不二家, 商品

表 3 複合語・構成語の分配トピック

	該当トピック番号
チョコまみれ	8, 21, 25, 76, 79
チョコ	8, 21, 25, 76, 79
まみれ	1, 3, 10, 12, 13, 15, 17, 18, 19, 20, 21, 22, 23, 25, 27, 31, 32, 33, 34, 35, 36, 38, 40, 41, 43, 54, 57, 58, 60, 61, 63, 68, 76, 79, 81, 83, 84, 85, 87, 89, 90, 92, 93, 94, 95

## 4 議論

### 4.1 まとめ

本研究では新規な商品名の複合語を取得し、その形式・意味に対して新規性・共通性を Google Trend 人気度の推移により分析した。その結果として、形式においては、構成語が新商品が現れる前から使われている共通性、複合語は新商品登場後に使われるようになる新規性があることが確認された。しかし多くの新規複合語商品名は調査を行った全期間で人気度が 0 であった。

意味の共通性に関しては、共起語分布の偏りで分析したが、分析対象とした複合語、構成語、複合語の上位カテゴリ語で大きな差は見られなかった。意味の新規性については、語が結合することで上位カテゴリ等以外の語が出現することが確認でき、構成語には無いイメージを複合語が持っている例があった。LDA による意味検出では、複合語のみに分配されたトピックを確認できなかった。

形式分析で新規性が評価できなかった理由として、ヒット商品の商品名だけを集めることの困難さが挙げられる。ある商標が一般消費者に広く使われるためには商標取得後に商品として販売し、その商品が広く普及することが必要になる。多くの商品は販売・普及の前に開発・販売が中止になることがあり、販売された中でヒットする商品はごく一部となる。今回の分析では、普及していない商品名も取得したため、このような結果になったと考えられる。その一方で、当初定義した新規性についての計測は困難だったが、「共通性がある語を結合し、今までにない形式の複合語を生成している」とはいえる。実際には使われていないため、新語とは言い難いが、もしその商品が普及し新語となった場合には仮説を肯定する語となることが予想される。

今回使用した GoogleTrend 人気度は、取得した期間における最大値を 100 とした相対的な値である。そのため、語の出現について絶対的な値を出していないため、出現の有無を言いにくい数値となっている。

LDA が抱える限界として、限られた意味のまとまりのものを多数のトピックに分解するとき、うまく分割できない点が挙げられる。トピック数を増やした分析をする必要がある。

## 謝辞

本研究は JSPS 科研費 JP 17H06383 の助成を受けた  
ものです。

## 参考文献

1. Lehrer, Adrienne. Understanding trendy neologisms. *Italian Journal of Linguistics*. 15: 369-382, 2003.
2. Shannon, Claude Elwood. A mathematical theory of communication. *The Bell System Technical Journal*, 27.3: 379-423, 1948.
3. Camiciottoli, Belinda Crawford. “All those Elvis-meets-golf-player looks”: A corpus-assisted analysis of creative compounds in fashion blogging. *Discourse, Context & Media*, 12: 77-86, 2016.
4. Lehrer, Adrienne. Neologisms. In: K. Brown (Ed.) *Encyclopedia of Language & Linguistics*. Second Edition, 590-593. Amsterdam: Elsevier, 2006.
5. 野田大志. 現代日本語における複合語の意味形成—構文理論によるアプローチ—. 名古屋大学大学院 国際言語文化研究科博士論文, 2011.
6. Fauconnier, Gilles, Turner, Mark B. Blending as a Central Process of Grammar, In: A. Goldberg, (Ed.) *Conceptual Structure, Discourse, and Language*. Stanford: Center for the Study of Language and Information, pp. 113-130, 1996.
7. Grady, Joseph E., Oakley, Todd, Coulson, Seana. Blending and metaphor, In: G. Steen & R. Gibbs (Eds.). *Metaphor in Cognitive Linguistics*, Philadelphia: John Benjamins, 1999.
8. 吉村公宏. 認知意味論の方法: 経験と動機の言語学. 人文書院, 1995.