

国立国語研究所学術情報リポジトリ

A Corpus-based Approach to Japanese Noun Noun Compounds

メタデータ	言語: jpn 出版者: 公開日: 2021-03-05 キーワード (Ja): キーワード (En): 作成者: 近, 大志, 神原, 一帆, Chika, Taishi, Kambara, Kazuho メールアドレス: 所属:
URL	https://doi.org/10.15084/00003142

日本語複合名詞へのコーパス基盤アプローチ

近 大志 (京都大学 大学院) *

神原一帆 (京都大学 大学院) †

A Corpus-based Approach to Japanese Noun Noun Compounds

Taishi Chika (Kyoto University)

Kazuho Kambara (Kyoto University)

要旨

複合名詞 (NN-Compound; NNC) は構成要素に還元できないような名詞間の意味関係を含むため、構成性を前提としたアプローチに対して重大な問題を喚起する (cf. Partee 1995). Levin et al. (2019) は「後部要素 N2 が自然種であるか人工物であるかによって、NNC の解釈のパターンが特定のクラスに収束する」という一般性の高い仮説を提唱した。本研究ではこの仮説が日本語においても有効かどうかを検討するため、BCCWJ から [格助詞 | 普通名詞 | 普通名詞 | 格助詞] のパターンに合致する事例を無作為に 500 件抽出し、N2 が人工物・自然物であるかを判定した。その結果、N2 は人工物・自然種の区別では説明できない要素を多分に含むだけでなく、この区分に該当する半数近くの事例が Levin らの説明からは十分に記述できない意味関係を持つことが判明したため、本仮説は日本語の NNC では必ずしも妥当ではないと結論づけた。

1. はじめに

複合名詞 (Noun-Noun Compound; NNC)⁽¹⁾ は名詞間の関係が構成要素に還元できない点で、意味解釈に関する議論が盛んである (cf. Downing 1977, Partee 1995, 影山 1999)。例えば、coffee cake には少なくとも「コーヒー風味のケーキ (i.e. コーヒーを**原材料とする**ケーキ)」や「コーヒーを飲む**際に**食べるケーキ」といった解釈が想定できるが⁽²⁾、**〈原材料とする〉**や**〈際に〉** という名詞間の意味関係は構成要素には存在しない⁽³⁾。

構成性を前提とした理論の限界を克服するため、構成要素のパターンを特定の意味関係に対応づけるアプローチが提案されてきたが、具体的な意味クラス (e.g., [N+ 事務用品], [N+ 食品]) に特化した議論が多い一方で、一般化を志向した研究は少ない (cf. Downing 1977, Ryder 1994, Wisniewski and Love 1998, Smith et al. 2014)。これに対して、Levin et al. (2019) は

* alberto_balsalm@aol.com

† kazy3024@gmail.com

(1) 本研究が主に扱う NNC は、前部要素 N1 が修飾部 (modifier) であり、後部要素 N2 が主要部 (head) であるような内心複合語 (endocentric compound) である。

(2) https://en.wikipedia.org/wiki/Coffee_cake

(3) 以降、N1 と N2 の意味関係を太字の **〈 〉** 内に表記する。

「後部要素 N2 が自然種であるか人工物であるかによって、解釈のパターンは特定のクラスに収束する」とする仮説 **Event vs. Essential Hypothesis (EEH)** を提示することで、英語における NNC の解釈のパターンが十分に予測づけられると主張した。EEH は記述的一般性が高いだけでなく、後述する理由から、通言語的にも成立することが想定されるため、この仮説が妥当であれば NNC の意味研究は大きく発展する。

本研究の目的は、日本語の NNC の分析を用いて Levin et al. (2019) の仮説 EEH を検証することである。BCCWJ を用いたコーパス調査の結果、(i) N2 が自然種・人工物に合致したとしても、Levin らの指定した意味関係では記述できない事例が半数近く含まれることに加え、(ii) NNC のうち、約 6 割が自然種・人工物以外を指示するため、EEH では NNC の意味解釈を十分に説明できないことの 2 点が判明した。これらの結果から、EEH が広範な NNC を説明できるという期待に反して、対象言語の相違を踏まえたとしても EEH の有効性が非常に限定的であると結論づけた。

本稿は次のように構成される。§2 では、構成要素のパターンに基づく NNC の意味解釈理論を概観し、本研究のリサーチクエスチョンを提示する。§3 と §4 では、BCCWJ を基に行った NNC の調査方法とその結果を報告する。§5 では結果に関する考察を行うと共に、Levin らの予測に合致しない事例について検討する。§6 は結論である。

2. 先行研究

本節では、本研究の検討対象である Levin et al. (2019) の論を中心に、NNC の意味解釈に関する先行研究を概観した上で、本研究のリサーチクエスチョンを提示する。

先述したように、構成性に基づく従来の意味解釈理論では NNC の意味を説明できないことが指摘されている (cf. Partee 1995)。Partee によると NNC の意味は、(i) 構成要素の意味に還元できない意味関係を含み (e.g., *a dog house* ⇒ *a house that a dog lives in*), (ii) 同じ要素から構成される NNC の意味関係がコンテキストによって変動する (e.g., *a dog house* ⇒ *a house that looks like a dog*). これらの理由から NNC は構成性を満たさないとされる。

(ii) に挙げた NNC の解釈の文脈依存性は主に語用論が扱う問題であるのに対して、(i) に挙げた意味関係の分析に関しては、構成要素の意味的なパターンを特定の意味関係と対応づける試みが知られている (cf. Downing 1977, Ryder 1994, Bezuidenhout 2019)。Wisniewski and Love (1998) は [N+ 事務用品] というパターンから 〈～に使用する〉 という意味関係が呼び出される傾向があることを心理実験によって観察した。それに加え、Smith et al. (2014) は [N+ 食品] というデンマーク語のパターンが 〈～を原材料とする〉 や 〈～を産地とする〉 という意味関係を喚起することを心理実験によって観察した。

構成要素のパターンと意味関係を対応づける一連の研究は、ある程度の成功を収めているとはいえ、適用範囲が比較的下位の概念 (i.e. 事務用品, 食品) に留まるため、NNC 一般に適用できる仮説が提供できない。その一方で、Levin et al. (2019) は [N+ 自然物] [N+ 人工物] といった上位概念から意味関係の予測づけを試みることによって、(1) に示す一般性の高い仮説を提示している。

(1) EVENTS VS. ESSENCES HYPOTHESIS:

人工物 (artifacts) を表す複合名詞は、自然物 (natural kinds) を表す複合名詞とは異なる傾向を示す。人工物では修飾語が人工物に関連する事象を指示する傾向がある一方、自然物では修飾語は自然物の本質を反映した性質を指示する傾向にある。

(Levin et al. (2019: 438); 筆者訳)

さらに Levin らは、自然物として解釈された NNC の意味関係および、人工物の意味関係に対する予測も与えており、これらは表 1 と表 2 から示される⁽⁴⁾。

表 1 N2=自然種の Meta Relation, 包含される意味関係, NNC の例

Meta Relation	包含される意味関係	例
〈借用〉	〈借用〉	keshi pearl
〈知覚〉	〈色〉〈大きさ〉 〈弁別的な部分〉〈風味〉 〈見た目〉	green bean, deep spoon, red-leaf lettuce, sugar pea, kidney bean
〈環境〉	〈産地〉〈場所〉	freshwater pearl, Boston lettuce

表 2 N2=人工物の Meta Relation, 包含される意味関係, NNC の例

人工物	包含される意味関係	例
〈事象〉	〈素材〉〈方法〉 〈目的〉〈時間〉 〈使用〉〈事態の担い手〉	opal bracelet, skillet cake, fish spatula, wedding band, navy bean, banana slicer

Levin らは 2 つの意味クラスに該当するパターン (自然種={ [N+ 鉱物], [N+ 植物] }; 人工物 = { [N+ 調理器具], [N+ 料理] }) のコーパスを作成し、人手によるコーディングを行った結果、NNC が EEH に従う傾向があることを観察した⁽⁵⁾。Levin らは非常に興味深い結果を示している一方で、仮説の一般性について 2 点の疑問が残る。

まず、EEH が英語以外の言語にも適用されるかどうかは議論の余地がある。Levin et al. (2019: 430) は Downing (1977) の論を援用することで、NNC を使う動機が事物の名付け (naming) に基づくと主張している。森岡・山口 (1985) は、日本語における名付けの手段の一つとして NNC を挙げている。このように、NNC を用いた名付けの行為は英語以外の言語にも観察されることから、EEH は名付けの一般的な傾向についての仮説として解釈することも可能であろう。この解釈が正しいと仮定した場合、英語以外の言語でも EEH が成立するか否かは検討すべき事項である。

⁽⁴⁾ 仮説の評価にあたっては、各意味関係を包含する Meta Relation が用いられている。また、Levin らは自然物の意味関係を予測づける仮説を指して Essence-Related Modifier Hypothesis と呼び、人工物の意味関係に関する仮説を Event-Related Modifier Hypothesis と呼んでいるが、ここでの議論には必要がないため省略する。

⁽⁵⁾ コーパスの構築は人手で行われたものであり、本論文執筆時点では <https://osf.io/t43kd/> に掲載されている。また、Levin らはコーパス調査だけでなく、理解・産出に関する実験を実施し、どちらの結果も EEH を支持すると結論づけた。

次に、EEH が調理器具や鉱物といった特定の意味クラス以外の語を含む NNC に対して有効か否かにも議論の余地がある。Levin et al. (2019) は [N+ 調理器具] や [N+ 鉱物] といった比較的具体的な対象の分析を通して EEH を提示しているが、EEH が高い一般性を志向している以上、NNC の解釈について過大般化を行なっている可能性が否定できない。よって、EEH の負例がどれだけ存在するかを調査する必要がある。

以上の2点を踏まえ、本研究では以下の2点をリサーチクエスチョンとする。

- (2) a. EEH の予測が日本語の NNC において再現されるか?
- b. EEH は NNC 全体を代表するものか?

3. 方法

前節でも述べたように、本研究の目的は Levin et al. (2019) の分析結果の一部を再現することである。本節では本研究が採用した調査方法について、データの収集方法、そしてそのデータの分析手法という順序でそれぞれ述べる。

本研究では現代日本語書き言葉均衡コーパス (BCCWJ) のデータを、中納言を用いて収集した。複合名詞の事例を収集するために短単位検索にて品詞情報を用いた検索を行った。可能な限り機械的に負例を取り除くため、「助詞-格助詞：名詞-普通名詞-一般：名詞-普通名詞-一般：助詞-格助詞」という検索条件を用いた。その結果 174,948 件のデータが得られた。そして、中納言を用いて収集した 10 万件をローカル環境に保存し、人手によるコーディングを行うために Ruby を用いて 500 件を無作為に抽出し、これを最終的な分析対象とした。

分析対象としたデータには、パーキングミスや従来の NNC に関する議論では分析対象として扱われないものを含むため、(3) に示されるデータを除外してある。特に、(広義の) 空間直示表現は英語などでは形容詞として用いられるため、NNC の議論には無関係であると判断した。

- (3) a. **全体が固有名を表す場合**: ワームテール, ロイヤルホテル, 達磨横町, 朝日新聞, 栃の木峠, ……
- b. **(広義の) 空間直示表現を含む場合**: 東西南北, 右手首, 中心シグナル, 中西部ツアー, 県境付近, ……
- c. **N1 または N2 が単純語ではない場合**: 同世代, 老若男女, 主素材, 学校側, オイル入り, ……
- d. **パーキングミス**: 装束ならび, 初句切れ, 楼閣机上, いきアレルギー, ……

(3) に挙げた負例は合計で 171 件となった。これらの事例を除いた 329 件の事例に対して、第 1 著者と第 2 著者は N2 が自然物か人工物のどちらを表すのかをコーディングし、EEH の傾向に従うか否かを Levin et al. (2019) のコーディング情報を参照しつつ判定した。

4. 結果

本節では、§3 で得たデータの調査結果を報告すると共に、EEH に従う事例および従わない例を分類する。表 3 に N2 の各意味クラスの集計結果を記載している。また、表 4 には、表 3 で自然種・人工物として判定された NNC が EEH に従うか否かの分析結果を掲載している。

各表の () にはそれぞれの値の相対頻度を記載している。

表3 N2 の分類結果

	自然種	人工物	その他	合計
N2	21 (0.064)	115 (0.350)	193 (0.587)	329 (1)

表4 EEH による分析結果

	正例	負例	合計
自然種	8 (0.381)	13 (0.619)	21 (1)
人工物	60 (0.522)	55 (0.478)	115 (1)

表3に示すように、Levinらの自然種と人工物という対立は必ずしも有効な分類にならないことが示唆される。それに加えて、表4の正例の列が示すように、EEHの予測は自然種であれば約4割が、人工物であれば5割程度しか適用されないことが判明した。以下では「その他」として分類された表現の一部、EEHの予測に従う事例、およびEEHの予測に従わない事例を概略的に挙げる。

はじめに、表3で「その他」に該当した名詞について議論する。§3で述べた方法で収集したデータの中には「夫婦」のような関係を表す語、「申し立て」や「会」といった事態を表す語、「愛人」のような、交際関係という関係の内の特定の役割を表す役割語が含まれる。これらはいずれも特定の状況との関係から特徴付けられる語であるため、自然種か人工物かという区分は有効ではないと判断した。それに加え、「効率」のようなある事柄の属性を表す表現や、「団体」や「全体」といったある対象の集合を表す語も、同様の理由から「その他」に分類した。

次に、表4で正例に該当した、EEHに従う事例について概観する。(4)はN2に自然種が当てはまり、(5)にはN2に人工物が該当する。

- (4) a. **〈知覚〉**：機銃音，自然環境，座標空間
- b. **〈場所〉**：我家玉ねぎ，イスラム世界，空中窒素，山津波
- (5) a. **〈素材〉**：りん酸ソーダ，ビニールクロス，紙ナプキン，……
- b. **〈目的〉**：メモリーカード，通販カタログ，道具箱，宗教法，……
- c. **〈使用〉**：警察車両，刑事部屋，男女言葉
- d. **〈事態の担い手〉**：オイルクーラー，水槽ヒーター，サウンドドライバー

(4)と(5)はどちらも、表1と表2のMeta Relationから記述が可能な例であるが、N1を事態として解釈するメトニミーを許容するのであれば、以下の例も人工物の**〈目的〉**として扱うことができる。

- (6) **〈目的〉(拡張版)**：下着ブランド，ラブホテル，水道メーター，テレホンカード，金券ショップ，ビームチューブ

そして、表4で負例に該当した、EEHに従わない事例について概観する。(7)は負例である自然物に当てはまり、(8)は負例の人工物に該当する。

- (7) a. **〈職業〉**：ベルボーイ
- b. **〈所有関係〉**：国有林野，拳銃男
- c. **〈使用用途〉**：公共空間，肉用子牛，……
- d. **〈構成要素〉**：レーザー幽霊，化学物質，……
- e. **〈その他〉**：少年少女，専門知識，化学物質，構造理性，保税地域
- (8) a. **〈場所の指定〉**：ローキャビネット
- b. **〈知覚的特徴の指定〉**：黒文字，縦文字，五輪塔，……
- c. **〈修辭的拡張〉**：大陸だな，赤ちょうちん，黒ミサ，……

最後に，Levin らが挙げていない意味関係のうち，Meta Relation の大幅な拡張によって記述が可能となるような事例を挙げる。

- (9) 音楽記事，歩兵学校，窓ガラス，人間像，……

5. 考察

本節では (2) に示したりサーチクエスチョンへの回答を与え，EEH の適用範囲が限定的であるということ指摘する。そして，日本語の NNC の分析を行うに際して (i) どのようなタイプの複合名詞が NNC として判定されるべきかを明示し，(ii) Levin らが採用した自然種／人工物という区分よりも詳細な意味クラスを措定する必要があることを述べる。

はじめに，(2a) の「EEH の予測が日本語の NNC においても再現されるか？」というサーチクエスチョンに回答を与える。

- (10) 表 4 が示すように，EEH の予測の有効性は限定的である。よって，EEH は日本語の NNC では再現されない。

自然種で EEH に従う事例は 4 割弱であり，人工物については EEH に従う事例の割合は若干増加するものの，これも 5 割程度に留まる。全体としてのサンプル数の少なさを度外視したとしても，表 4 の結果からは，EEH の予測がいずれもチャンスレベルでしか支持されない可能性が示唆される。

次に，(2b) の「EEH は NNC 全体を代表するものか？」というサーチクエスチョンに回答を与える。

- (11) 表 3 が示すように，6 割程度の NNC が自然種・人工物以外を表すため，EEH が NNC 全体を代表するものとは断定できない。

ただし，人工物に関しては，その対象が参与しうるような事態を表しうるという点において改善の余地が見られる。例えば，(9) に挙げた「音楽記事」などは執筆活動と執筆内容から成る関係を想定することで EEH でも説明できると考えられる。同様に，「歩兵学校」についても養成活動と養成の結果状態から成る事態を想定することで説明が可能になると考えられる。しかし，このような事例に逐一特定の事態を割り当てることはアドホックな説明に過ぎず，理論の根本的な見直しが必要となる。

以上を踏まえると，日本語の NNC の分析には，どのようなタイプの複合名詞が NNC として判定されるべきかを明示した上で，Levin らが採用した自然種／人工物という区分よりも詳細な意味クラスを措定する必要があると考えられる。これらについて以降順次に議論する。

まず、日本語における NNC の同定法に関する問題について考察する。一般的に NNC の分析で問題となるのは *dog* や *house* といった具体的な指示対象を持つ語の組み合わせであった。しかし、より妥当な複合名詞の組み合わせとしては (3a-c) に挙げたような事例を体系的に排除する必要がある。特に「右手首」のような事例は英訳すれば *the right wrist* と形容詞を含む名詞句として表現されるため、英語で議論されていたような NNC の分析が日本語に直接適用できるかは疑わしい。

(3c) で挙げたような事例は N1 または N2 を独立した名詞として扱うことが難しいことを上では指摘した (e.g., 同世代)。このような事例は [同 N] という形式でしか用いられないものであり、一種の接辞として機能する。これに対して (4) と (5) で示した事例に含まれる名詞 (e.g., 我家玉ねぎ, リン酸ソーダ) は統語論的に様々な位置に生起することができる。これら**タイプ頻度** (type frequency) の差は一般的に**生産性** (productivity) の問題として解釈することもできる。よって、日本語における NNC の同定にこれらの生産性の差を加味することは今後重要になると考えられる。

次に、Levin らは NNC の解釈に自然種／人工物という区分が有用であると繰り返し強調するが、この二分法が全ての名詞にとって適切な分類を提供するとは限らないことを §4 にて指摘した。両者の明確な区別が設けられない事例があるという点は Levin らも認めるものの⁽⁶⁾、特に何らかの対象間の関係を表す事態名詞 (e.g., 三墨ゴロ) や、ある状況の役割を表すような役割名詞 (e.g., 女性兵士) の扱いについては手付かずの状態である。これらの語を EEH で扱うためには、**フレーム意味論** (frame semantics) のような理論的枠組みを採用することが解決策の 1 つとして望まれる (cf. Fillmore and Baker 2015)。しかし既に EEH の拡張について指摘したように、各事例に対して逐一特定の事態を割り当てることはアドホックな説明に陥る可能性があるため、体系的なタグ付けに関する理論的な考察を要する。

ただし、フレーム意味論を補助仮説として採用したとしても「経済社会」や「産業構造」, 「賃金格差」や「オゾン層」, 「宗教団体」といった語の扱いに関してはいまだに議論の余地がある。NNC の解釈に自然種と人工物以外の意味クラスも重要な役割を果たすということ自体は疑いようがないと思われる。しかし、アドホックな解決を避けるためにも無闇に意味クラスを増やすことには慎重になった方が良好だろう。

6. おわりに

本研究では、構成要素の意味的なクラスから NNC の解釈の傾向性を捉えるアプローチを要約し、「後部要素 N2 が自然種であるか人工物であるかによって、解釈のパターンが特定のクラスに収束する」とする Levin et al. (2019) の仮説 EEH を検討した。コーパス分析の結果、EEH における自然種／人工物という区分は、必ずしも NNC の解釈に対して適切な予測づけを提供するとは断定できないことが示された。そして、EEH による分析が困難である原因として (i) 日本語の NNC は必ずしも英語で扱われている NNC とは対応しないことに加え、 (ii) EEH

⁽⁶⁾ Levin et al. (2019: 438) は自然種と人工物の線引きが難しい例の 1 つとして *cart horse* を挙げている。ウマ自体は自然種であるが、*cart horse* が指示するのは馬車を牽引するために使役されるウマであり、特定の役割を担っている点で人工物として解釈される。

による分析を拡張するには、名詞の意味クラスの整備が必須となることの2点を主張した。

名詞の意味クラスの同定は、シソーラスといった言語資源の開発には欠かせないが、具体的にどのような名詞の分類が有効になりうるのかという点については未検討に終わった。また、EEHを補完する理論の1つとして提案したフレーム意味論がどの程度EEHと整合するのか、という点についても多くの議論を必要とする。これらに関しては今後の課題としたい。

文 献

- Pamela Downing (1977). “On the creation and use of English compound nouns.” *Language*, 53:4, pp. 810–842.
- Barbara Partee (1995). “Lexical semantics and compositionality.” Lila Gleitman, and Mark Liberman (Eds.), *An Invitation to Cognitive Science: Language*. Vol. 1.: MIT Press. pp. 311–360.
- 影山太郎 (1999). 『形態論と意味』 くろしお出版.
- Mary E. Ryder (1994). *Ordered Chaos: The Interpretation of English Noun-Noun Compounds*.: University of California Press.
- Edward J. Wisniewski, and Bradley C. Love (1998). “Relations versus properties in conceptual combination.” *Journal of Memory and Language*, 38:2, pp. 177–202.
- Viktor Smith, Daniel Barratt, and Jordan Zlatev (2014). “Unpacking noun-noun compounds: Interpreting novel and conventional foodnames in isolation and on food labels.” *Cognitive Linguistics*, 25:1, pp. 99–147.
- Beth Levin, Lelia Glass, and Dan Jurafsky (2019). “Systematicity in the semantics of noun compounds: The role of artifacts vs. natural kinds.” *Linguistics*, 57:3, pp. 429–471.
- Anne Bezuidenhout (2019). “Noun–noun compounds from the perspective of Relevance Theory.” Kate Scott, Billy Clark, and Robyn Carston (Eds.), *Relevance, Pragmatics and Interpretation*.: Cambridge University Press. pp. 174–186.
- 森岡健二・山口仲美 (1985). 『命名の言語学ネーミングの諸相』 東海大学出版会.
- Charles J. Fillmore, and Collin F. Baker (2015). “A frames approach to semantic analysis.” Bernd Hein, and Heiko Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press. pp. 791–816.

関連 URL

コーパス検索アプリケーション『中納言』 <https://chunagon.ninjal.ac.jp/>