

名詞辞書におけるコロケーションの記述

3G-6

桑畑和佳子*, 橋本三奈子*, 井口厚夫**, 猪塚元***, 村田賢一*

*情報処理振興事業協会(IPA) **独協大学 ***東邦大学

1. はじめに

IPAL名詞辞書では、名詞のコロケーションの記述の試みの一つとして、その名詞と結びつく述語を名詞の〈意味素性〉別に収録している。以下、これを「コロケーション」と呼ぶ。〈意味素性〉とは、結びつく述語によって焦点が当てられる名詞の意味的な側面を表すものである。例えば、〈具体物〉という側面で「車」という名詞を捉える述語には「車が壊れる」「車を磨く」があり、〈自動的に動くもの〉という側面では「車が走る」「車で行く」などがある。

これらの述語は主に新聞・教科書・論説文・小説類などおよそ43万文のコーパスを分析して抽出したものである。しかし、単純にコーパスを分析して得られるものだけではコロケーションの記述として不十分であることが多い。自然に結びつくものであるとは言いがたいものがコーパスに共起出現している場合（例：「皿で殴る」）や、見出し語に共起すると考えられる述語がたまたまコーパスに現れていない場合（例：「食堂が混む」）があるからである。前者の場合には、抽出する際に取り除いた。後者の場合には、共起可能性をチェックして網羅的に収録するものと、可能性はあっても収録しないものとに述語のタイプで仕分けた。本論文では最初に、コーパスから抽出できるコロケーションの例を示し、続いてコーパスに現れにくいコロケーションを述語のタイプによって収録仕分けた方針について述べる。

On lexical description of nominal collocations

Wakako Kuwahata, et al.,

Information-technology Promotion Agency

2. 見出し語に特有の述語（〔文献1〕）

以下の述語は見出し語に特有であるため、コーパスから抽出することで網羅的に収録できる。

(1) 構文的な役割を担うもの

【選挙】がある 【伝統】がない
【発表】をする 【逆さま】にする
【孤独】になる

(2) 見出し語固有で他の述語と言い換え可能なもの

【判定】を下す=判定する、判定をする
【反対】にあう=反対される、反対を受ける

(3) 見出し語固有で出現頻度が高いもの

【ズボン】をはく | を脱ぐ | が長い
【電気】をつける | を消す | が暗い

3. 見出し語にある程度特徴的な述語

(1) 出現頻度は低いが、結びつく名詞に制限があるもの

一般の国語辞典で「切らす」をひくと、用例としてあがっているのは「たばこを切らす」だけ、あるいは「小銭を切らす」までの二例であった。コーパスでは「潤滑油を切らす」「ガソリンを切らす」の二例が見つかった。しかし、上の他にも「塩」「酒」「えさ」「くすり」「食料」など食べ物を表すいくつかの名詞は「切らす」という述語と結びつくことができる。但し、同じ食べ物を表す名詞でも、「蕎麦を切らす」「焼き肉を切らす」などは言いにくい。食べ物以外の具体物名詞では「紙」「花」「石鹼」などは「切らす」と結びつくが、「机」「靴」「タンズ」が「切らす」に結びつくことは考えられない。以上のことから

「切らす」という動詞と結びつくことが可能な名詞は「ストックするもの」であるといえそうである。このように、結びつくつかないかによって名詞のある側面を浮き彫りにさせることのできる述語は網羅的に収録すべきであるので、関連する名詞全てについて共起の可否をチェックしている。

(2)見出し語がその〈意味素性〉をもつと判断できるような典型的なもの

「蕎麦」という名詞を〈食べ物〉と捉える時の述語欄には見出し語特有の「すする」の他、〈食べ物〉として捉える根拠となる「食べる」「味わう」といった述語も収録した。これらの述語は、〈意味素性〉ごとに共通につき得やすいものとしてリストアップすることができる。このリストは似たような見出し語について共起の可否をチェックするのに役立つ。他に「つまむ、吐く、うまい、まずい、甘い、からい」などが挙げられる。「蕎麦を食べる」は一般のコロケーションの概念としてはかなり結びつきが緩いものではあるが、「スープを食べる」とは言わないといった場合もあり、網羅的な収録に価値があるといえる。

4. 結びつく名詞に強い制限のない述語

さて、3.の態度とは反して、単に「結びつくことが可能である述語」については網羅的に記述していない。網羅的に記述しようと試みると、例えば「蕎麦」という名詞を〈具体物〉と捉える時には「投げる」「捨てる」「上げる」「見る」「触る」「とる」…などの述語も収録していくということになり、きりがない。否定表現での結びつきも許すならさらに共起可能な述語が増えていくことになり、この作業には無理があることがわかる。よって、「蕎麦」の場合には「投げる」「捨てる」などの述語は収録しないことにした。但し、このような「結びつく名詞に強い制限のない述語」は「ボールを投げる」「ゴミを捨てる」のようにコ

ーパスにおいて出現頻度の高かった名詞にはその述語欄に例として収録することにした。

このようなタイプの述語としては他に「欲しい」がある。コーパスでは、「時間」「金」「仕事」「自動車」などの名詞は、「欲しい」という形容詞と共起している。しかし、この他「庭」や「くつ」なども「欲しい」と結びつくことは可能であるがコーパスには現れていない。コーパスが大きくなればもっと共起出現する名詞が増えそうであるが、だからといって「塩」「しょうゆ」「鍋」などの名詞にまで「欲しい」を収録するのは、冗長であると感じられる。IPAL形容詞辞書には「ほしい」のとり名詞句は、〈具体物〉でも〈抽象物〉でもよく、「その情報、あの本、水、おやつ、いいカメラ、このスカート、名誉、人材」と例があがっている。従って、コーパスに共起出現していた「時間」「金」「仕事」「自動車」などの述語欄には「欲しい」を収録したが、出現していなかった名詞に網羅的に収録するということはしなかった。

5. おわりに

IPAL名詞辞書に収録したコロケーションはコーパスから単純に抽出しただけの結果ではない。その述語が「見出し語に特有のもの」であればコーパスから抽出し、「見出し語に特徴的なもの」と考えられるものはコーパスと内省によって網羅的なチェックを行い、「結びつく名詞に強い制限のないもの」と考えられるものはコーパスで出現頻度の高い名詞についてだけ抽出し、〈意味素性〉ごとに分類して収録したものである。

IPAL動詞辞書、形容詞辞書との連携については今後検討していく予定である。

参考文献

- 〈1〉橋本三奈子ほか(1993)「コーパスデータに基づく名詞コロケーションの辞書記述」『情報処理学会第47回全国大会論文集』3-63