

# 解説

## 計算機で言葉の意味を扱う難しさ — オノマトペのあいまいさを例に —†

内田 ゆず\*

### 1. はじめに

人工知能における難問の1つに記号接地問題（symbol grounding problem）がある [1]。これは、記号が指し示すものを計算機がどのように認識するかという問題である。現状では記号（言語表現）を現実世界と結びつける方法は確立されておらず、自然言語処理の分野では重要な課題である。

近年、深層学習の手法を自然言語処理のタスクに適用することで、有効な結果が得られている。自然言語を深層学習で扱うためには、単語の意味を低次元のベクトルで表現（分散表現）する必要がある。分散表現を構築する際には、ニューラルネットワークに基づく言語モデルが用いられることが多い。中でも、[2]で提案された CBoW と Skip-gram を用いた word2vec は広く使われている。word2vec は  $King - Man + Woman = Queen$  の例とともに紹介され、あたかも単語の意味を捉えているような演算ができるということでも話題になった。

単語の分散表現は自然言語処理の分野で大きく貢献しているが、問題点もある。その1つは「似た意味の単語は、似たような文脈で使用される」ことが前提となるモデルであるため、対義語や多義語を正確に処理できないことである。さらに根本的な問題として、分散表現は学習データの中（記号の世界）で完結しており、現実世界とのつながりがないことも挙げられる。このように、「ある程度はうまくいくが完璧ではない」状況は自然言語処理に限らず、人工知能技術全般に見られる。

現在の技術では言葉の意味の本質を機械で処理することはできない。“意味を計算する”方法が確立するまでは、語義を機械可読な形で地道に記述することも重要である。それらのデータは機械学習の精度向上にも寄与するだろう。

著者は、日本語の表現の中でも特にあいまいな語群であるオノマトペの語義について研究を続けている。本稿では、オノマトペの工学的応用を目的として著者がこれまでにやってきた研究について紹介し、言語がもつあいまいさを計算機で扱う方法を考える。

### 2. オノマトペのあいまいさ

オノマトペは、現実世界の音や形、様子、動き、状態を写し取った言葉であり、微妙なニュアンスを伝える際に役立つ。また、多義性が高い、表層の変形に寛容である、品詞が一意に決まらないなどの特徴をもつ。そのため、新語を作ることや、既存のオノマトペを新たな語義で使うことが容易にできる。このような豊かな表現力や用法の自由度から、オノマトペは日常の日本語に不可欠な語群である。一方で、計算機での処理を考えたとき、この柔軟さは大きな障壁となる。以下では、日本語オノマトペのあいまいさについてさまざまな角度から分析した例を紹介する。

#### 2.1 語義と用法の多様性

大部分のオノマトペは、複数の語義をもつ。たとえば、オノマトペ「ぶらぶら」には以下の4つの語義が定義されている [3]。

1. 宙に釣り下がって揺れ動くさま
2. あてもなく時間をかけて歩くさま
3. 何の目的もなく怠惰に暮らしているさま
4. 原因がはっきりしない病気が長引いて、なかなか治らないさま

また、オノマトペの品詞は一定せず、副詞、名詞、連体詞などとしてふるまう。例として「ぶらぶら」の

† Difficulty of Computing the Meaning of Words: A Case of Onomatopoeia Words  
Yuzu UCHIDA

\* 北海学園大学工学部電子情報工学科  
Faculty of Engineering, Department of Electronics and Information Engineering, Hokkai-Gakuen University

表 1 オノマトペの用法

|         |            |
|---------|------------|
| 副詞的用法   | 近所をぶらぶら歩く  |
| サ変名詞的用法 | 家でぶらぶらする   |
| 連体詞的用法  | ぶらぶらのイヤリング |
|         | ぶらぶらした腕    |

用法を表 1 に示す。

## 2.2 時間変化

オノマトペの歴史は古く、奈良時代には既に用いられていた [4]。最古の例は「古事記」に見ることができる。また、長い年月を経てオノマトペの語義や用法は変化することもわかっている [5]。

著者らは現代日本語におけるオノマトペの語義・用法の実態を捉えるために、地方議会会議録コーパス<sup>1</sup>に含まれるオノマトペを分析し、新しい用法を発見した [6]。例えば、オノマトペ辞典 [3] によると「ばくっ」は「大きな口を開けて勢いよく食いつくさま」などの意味をもつが、コーパス中では「漠然とした、大ざっぱな」の意味で用いられている。このことから、比較的短期間でも語義は変化するといえる。

## 2.3 地域差

オノマトペのあり方は地域によって異なり、方言学的見地からの研究もなされている [7-9]。

著者らによる地方議会会議録の分析によると、オノマトペの使用頻度が高いのは中国地方と近畿地方であった [6]。また、ある地域特有の用法も発見された。「ぴしゃっ」は九州地方で多く出現し、「きちんと」「しっかりと」に類する意味で使用されている。全国的には「ある基準で完全にやめる、閉める、止めるさま」の意味で用いられており、オノマトペの語義の地域差を示す結果となった。

## 2.4 ジャンル・文体による差異

オノマトペはカジュアルな文書で使用されることが多いが、一部の語はフォーマルな文書でも使用される。そこで、ジャンルによるオノマトペの用法の違いを分析した。会議録（フォーマル）と BCCWJ（カジュアル・フォーマル混合）を比較したところ、オノマトペおよび係り先動詞の出現傾向に偏りがみられた [10]。会議録では「ばっさり-削る」や「とんとん-なる」、BCCWJ では「のんびり-暮らす」などの出現頻度が高い。Web 文書（カジュアル）と会議録（フォー

マル）を比較した場合も、オノマトペの係り先動詞に異なる傾向があった [11]。このように、オノマトペの使われ方は文書のジャンルや話題に依存する。

## 3. オノマトペの工学的応用

本章では、オノマトペ研究の応用事例を紹介する。

### 3.1 自然言語処理の基礎技術

日本語を対象とした自然言語処理のタスクの多くは、形態素解析によって単語を抽出することから始まる。しかし、オノマトペを含む文は形態素解析誤りを起こしやすいことがわかっている [12]。これは、オノマトペがひらがなで記述される短い単語であることや、品詞が一意に定まらないことなどが原因である。

著者らは文書からオノマトペを含む文を自動抽出するアルゴリズムを構築した [13] が、未だ改善の余地がある。さらに、オノマトペの意味を正確に捉えるためにはオノマトペが修飾する対象を同定する必要がある。構文解析の精度向上も急務である。

### 3.2 日本語教育

日本語は特にオノマトペに富む言語であり、他言語には日本語のオノマトペに相当する語がない場合がある。また、オノマトペは多義性が高く、文脈によって異なる語義で使用される。したがって、オノマトペを含む日本語を他言語に翻訳することは困難である。

これは、日本語学習者にとっても避けては通れない問題である。オノマトペは他の語で言い換えることが難しく、辞書の定義だけでは正確なニュアンスを理解することができない。類似のオノマトペの使い分けも日本語学習者にとっては難題である。

著者らは、日本語学習者の協力を得て、オノマトペ学習に必要な要素を調査した。その結果、オノマトペの意味とともに動画や静止画といった視覚情報を提示すると、オノマトペの理解が促進することを明らかにした。これらの知見を利用してマルチメディアを導入したオノマトペ学習支援システムの開発も行っている [14, 15]。

記号（言語表現）をマルチメディア情報と結びつけることは、現実世界との接点を与えることだともいえる。今後は、種々の場面でのオノマトペの意味・用法を、動画、静止画、音声を活用して定義したい。

## 4. おわりに

本稿では、言葉のあいまいさについてオノマトペの分析結果を題材にして解説した。著者らは、第 2 章で述べた分析結果を活用し、オノマトペの用法およびコ

<sup>1</sup> 地方議会会議録コーパスプロジェクト  
<http://local-politics.jp/>

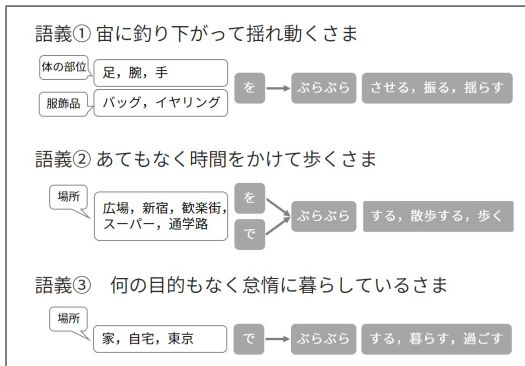


図 1 コロケーションと語義

ロケーションと語義の関係を分析している（図 1 参照）。さらに、これらの知見をオノマトベの語義曖昧性解消に応用し、オノマトベの意味の問題に取り組んでいる [16, 17]。

オノマトベのみならず言葉の意味全般を計算機で扱う技術が実現し、ソフトコンピューティング分野がますます発展することを期待したい。

## 謝辞

本研究の一部は科研費（若手研究 (B)23700256, 基盤研究 (C)17K00294, 若手研究 (B)17K12791）の助成を受けたものである。

## 参考文献

- [1] S. Harnad: "The Symbol Grounding Problem," *Physica D: Nonlinear Phenomena*, Vol. 42, pp. 335-346, 1990.
- [2] T. Mikolov, K. Chen, G. Corrado, and J. Dean: "Efficient Estimation of Word Representations in Vector Space," *Proc. of Workshop at ICLR*, 2013.
- [3] 小野正弘: "日本語オノマトベ辞典," 小学館, 2007.
- [4] 小野正弘: "感じる言葉 オノマトベ," KADOKAWA/角川学芸出版, 2015.
- [5] 中里理子: "オノマトベの語義変化研究," 勉誠出版, 2017.
- [6] 高丸圭一, 内田ゆず, 乙武北斗, 木村泰知: "地方議会会議録コーパスにおけるオノマトベ出現傾向と語義の分析," *人工知能学会論文誌*, Vol.30, No.1, pp. 306-318, 2015.
- [7] 竹田晃子: "円滑な医療コミュニケーションのための方言集: 『東北方言オノマトベ用例集』の取り組み," *日本方言研究会研究発表会発表原稿集* 94, pp. 55-58, 2012.
- [8] 方言研究ゼミナール: "身体感覚を表すオノマトベ," *方言資料叢刊*, 第 2 巻, 1992.
- [9] 小林隆編: "感性の方言学," ひつじ書房, 2018.
- [10] 高丸圭一, 内田ゆず, 乙武北斗, 木村泰知: "係り先動詞に着目したオノマトベの語義分類に関する検討," *知能と情報（日本知能情報ファジィ学会誌）*, Vol.28, No.4,

pp. 693-699, 2016.

- [11] 内田ゆず: "2 種のコーパスを対象とした係り先動詞のエントロピーに基づくオノマトベの特徴分析," 第 33 回ファジィシステムシンポジウム講演論文集, pp. 707-710, 2017.
- [12] 内田ゆず, 高丸圭一, 乙武北斗, 木村泰知: "BCCWJ コーデータにおけるオノマトベ出現実態の分析—現代オノマトベ実例辞書アプリ構築に向けて," 2015 年度人工知能学会全国大会, 3G4-OS-05b-5, 2015.
- [13] 内田ゆず, 荒木健治, 米山淳: "ブログ記事からのオノマトベ用例文の自動抽出手法," *知能と情報（日本知能情報ファジィ学会誌）*, Vol.24, No.3, pp. 811-820, 2012.
- [14] 上野駿太, 内田ゆず, 長谷川大, 佐久田博司: "エージェントによるオノマトベ教育システムの開発," *情報処理学会第 75 回全国大会講演論文集*, pp. 4661-4662, 2013.
- [15] 家坂光弥, 内田ゆず, 米山淳: "日本語学習者のオノマトベ学習を支援するタブレット端末用アプリケーションの開発と有効性の評価," 第 30 回ファジィシステムシンポジウム講演論文集, pp. 632-637, 2014.
- [16] H. Fukushima, K. Araki, and Y. Uchida: "Disambiguation of Japanese Onomatopoeias Using Nouns and Verbs," *Lecture Notes in Artificial Intelligence*, Vol.1, 8655, pp. 141-149, 2014.
- [17] 内田ゆず, 高丸圭一, 乙武北斗, 木村泰知: "BCCWJ におけるオノマトベの語義とコロケーションの分析," 2016 年度人工知能学会全国大会, 1K4-OS-06a-2, 2016.

(2018 年 6 月 15 日 受付)

[問い合わせ先]

〒064-0926 北海道札幌市中央区南 26 条西 11 丁目 1-1

北海学園大学工学部電子情報工学科

内田 ゆず

E-mail: yuzu@eli.hokkai-s-u.ac.jp

## 著者紹介



うちだ  
内田 ゆず [正会員]

2005 年北海道大学工学部情報工学科卒業。2010 年同大学院博士後期課程修了。博士（情報科学）。同年、青山学院大学理工学部助手。2012 年より同助教。2014 年北海学園大学工学部電子情報工学科准教授、現在に至る。自然言語処理、特に概念獲得システム、日本語オノマトベなどに関する研究に従事。日本知能情報ファジィ学会、電子情報通信学会、言語処理学会、人工知能学会、情報処理学会各会員。