



北海道大学  
HOKKAIDO UNIVERSITY

# オノマトペの多義性解消における名詞の有効性

北海道大学 大学院情報科学研究科<sup>†</sup>

北海学園大学<sup>‡</sup>

福島 弘識<sup>†</sup> , 内田ゆず<sup>‡</sup> , 荒木健治<sup>†</sup>



複数の意味をもつオノマトペが

文の中でどの意味で使われているかを

自動判別するシステムの構築

※オノマトペ…擬音語,擬態語の総称



# 背景

- ◆ 日本人はオノマトペを本当によく使う

→「サクサク♪お家でアメリカン\*アップルパイ」(cookpad<sup>1</sup>より)

→「赤字が出ようがぴしゃっとボーナスも出ておる」<sup>2</sup>(長崎県)

- ◆ 日本語の正確な理解にはオノマトペは不可欠

→機械翻訳や、ロボットの言語理解など

- ◆ オノマトペが文中でどのような意味で使われているか判別する必要性

→これは擬音？擬態？

→1語で複数の意味を持つオノマトペもたくさんある

1) Cookpad : <http://cookpad.com/>

2) 高丸圭一, 内田ゆず, 乙武北斗, 木村泰知: 地方議会会議録におけるオノマトペの出現傾向に関する基礎的検討 -少数の自治体に高頻度で出現するオノマトペについて-, 言語処理学会第20回年次大会, P5-16, 2014.



“喉がカラカラ”

すみません、よくわかりませ  
ん。

“喉が渴いた”

タップすると編集できます

このような場所なら、渴きを  
いやせるでしょう:

はい、見つかったのはこちら  
です:

15件の結果

この周辺



同じでもいいはずでは？

カラカラだけで理解できれば  
一流の日本語話者だ！



badly??

物がガタンと落ちた  
体調がガタンと落ちた  
電車がガタンと揺れた

- 英語
- 中国
- 韓国
- フランス

A thing fell off badly  
Physical condition fell off badly  
A train shook with ガタン

Gatan??

日本語

物がガタンと落ちた  
体調がガタンと落ちた  
電車がガタンと揺れた

英語

Things fell and Gatan  
Physical condition fell and Gatan

このように1語で複数の意味をもつオノマトペの

moth tongue 意味判別はまだ完成していない

英 日  言語自動判定

物がガタンと落ちた  
体調がガタンと落ちた

クリア  
クリア

日 英

A thing fell with moth tongue.  
Physical condition fell with moth tongue.

翻訳



形容詞？

名詞？

どうやって判別しよう？

動詞？

係り先？



# 本研究のアプローチ

- 1, オノマトペの意味を決定する単語を自動抽出する
- 2, オノマトペの意味を決定する単語のどんな性質がオノマトペの意味判別に有効か分析する



## システムの構築

入力: オノマトペ(1語で複数の意味をもつ)を含む文

出力: 文中に出現するオノマトペの意味



# 先行研究<sup>1)</sup>

—機械学習を用いた多義性判別の結果と分析—

精度[%]	名詞	係り先動詞	名詞+係り先動詞
“さらり”	72.7	85.2	83.4
“がたん”	80.7	81.7	82.5

- \* 係り先動詞と名詞を素性として一定の成果
- \* 形容詞の効果がないことも確認
- \* 研究対象としたオノマトペが少ない → 対象オノマトペを増やす必要性
- \* オノマトペ毎に多様な特徴や表現の出現 → 汎用性をもつ判別法の模索

1) Hironori Fukushima, Kenji Araki, and Yuzu Uchida :

"Disambiguation of Japanese Onomatopoeias Using Nouns and Verbs" P. Sojka et al. (Eds.) :

TSD2014, LNAI 8655, pp.141-149, 2014



オノマトペの種類を問わず、  
オノマトペの意味を判別できるような、  
出現単語(主に名詞)の使い方や  
ルールを設定を試みる。  
ルールと機械学習を組み合わせ  
判別システムの構築を目指す。



オノマトペの意味を決定する単語を抽出する



# オノマトペの意味を決定する単語の自動抽出

運動／**量**／が／ガタン／と／落ちる／後半

オノマトペの係り先動詞に係る名詞によって決定される場合

電車／が／ガタン／と／大きく／**揺れた**

オノマトペの係り先動詞によって意味が決定される場合

突然／ガタン／と／大きな／**物音**

“音”という単語が含まれることで  
必然的に擬音語と決定される場合



先行研究でオノマトペの係り先動詞を  
素性にした機械学習で、いい精度だったし…

オノマトペの係り先みたら  
だいたいできるでしょ？

いまの例にもオノマトペの係り先で  
決まってしまう例もあったし…





# オノマトペの係り先は直接意味決定に関係するか？

	オノマトペの係り先		
	動詞	名詞	その他
対象文数	164	18	5
抽出単語	52	16	0

抽出単語…被験者がオノマトペの意味を決定すると判断した単語

⇒ 係り先動詞のみでは意味を決定できないことが確認された



# オノマトペの意味を決定する単語の自動抽出

- ・ 係り先動詞のみで意味が決まる  
→ 動詞をリスト化
- ・ 係り先動詞のみで意味が決まらず名詞に依存する  
→ 名詞抽出のため、**詳細なルールを設定**



# ルールの一例

楽天は/岩隈と/田中は/素晴らしいのですが、

0 1 2 3

他の/先発**投手**だと/ガタンと/**力**が/落ちるんですよ。

4 5 6 7 8

候補語

投手

力

ガタン

落ちる

$$D_{on} = |N_{pos} - O_{pos}| \quad \text{どちらも 1}$$

$$D_{kn} = |N_{pos} - K_{pos}| \quad \text{“投手”} = 3 \quad \text{“力”} = 1$$

抽出名詞



オノマトペの意味を決定する名詞の**どんな性質**が

オノマトペの**意味判別**に有効か分析する



# 抽象物と具体物の判別

“ガタン”を含む文で分類した結果

- JUMAN のカテゴリ情報を用いた分類
- 分類対象の名詞は、オノマトペの意味を決定する名詞

抽出名詞

“力” = 抽象物

ガタン(擬態)と力が…

カテゴリ		擬音語	擬態語
抽象物	音	36	0
	音以外	0	45
人工物	乗り物	2	0
	金銭	0	3
	その他	5	3
場所		4	0
数量		0	21
時間		0	1
自然物		0	1
動物		0	1
カテゴリなし		0	7

\*擬音語か擬態語どちらかに偏っている



# 抽象物と具体物の判別

- ・ **抽象物**… ”抽象物”， ”色”， ”数量”， ”時間”， ”形・模様”， ”人工物一金銭”
- ・ **具体物**… それ以外の全てのカテゴリ

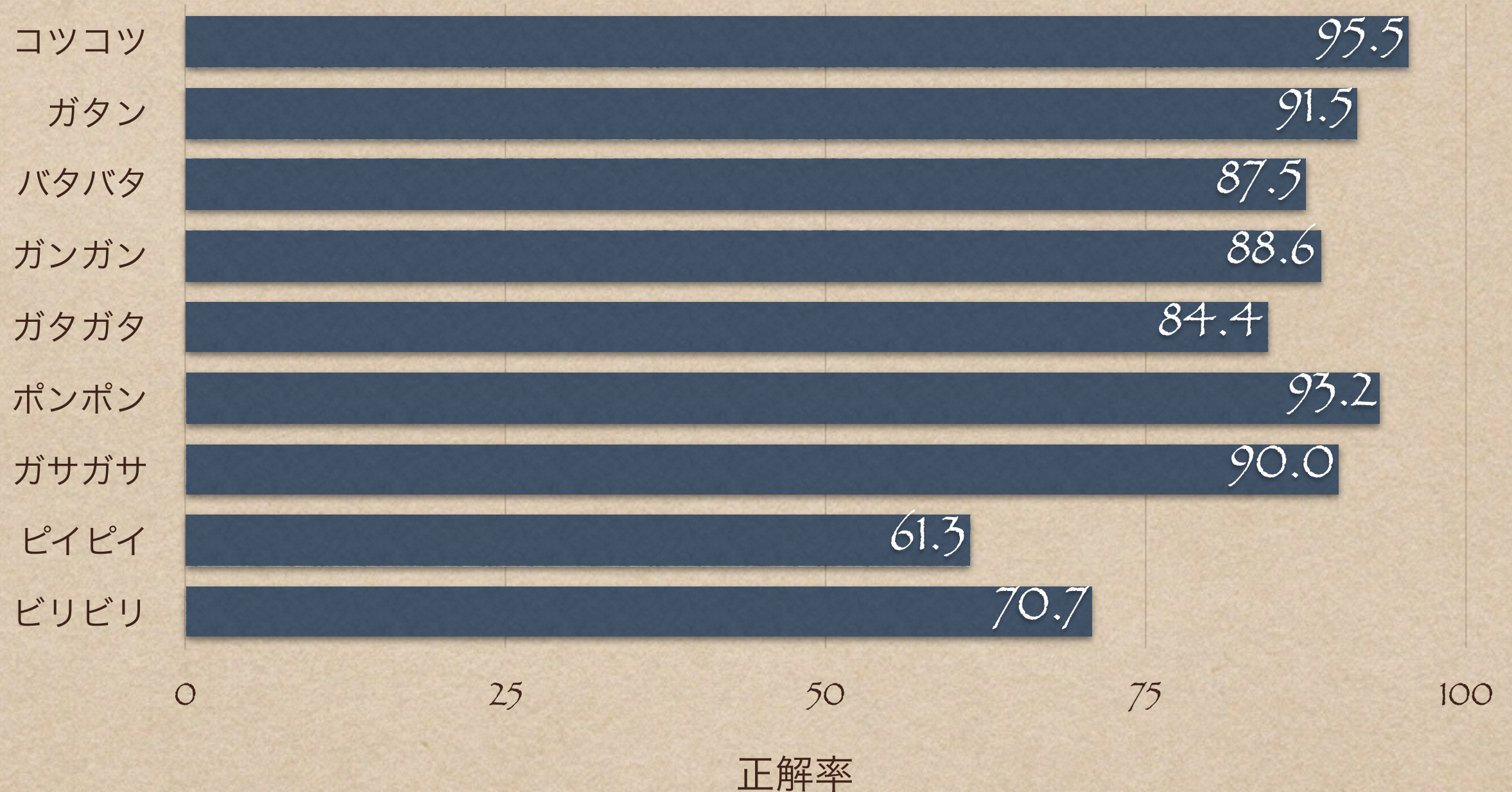
\* オノマトペの意味決定名詞(抽出名詞)が抽象物なら**擬態語**，  
具体物なら**擬音語**

\* 抽出名詞に”音”が含まれれば**擬音語**



# オノマトペ毎の意味判別正解率と考察

(名詞によって意味が判別される全450文の結果)



- ◆ 動詞によって意味が判別される文は非対象
- ◆ 平均正解率は**86.2[%]**



# オノマトペの意味を決定する名詞&動詞を用いた 意味判別システムの構築

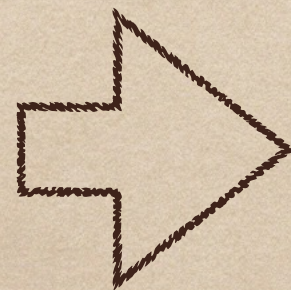
2:21  
ark4.media.eng.hokudai.ac.jp

DISAMBIGUATION SYSTEM  
for  
JAPANESE ONOMATOPOEIAS  
オノマトペ意味判別システム

!!現在判別可能なオノマトペの種類!!  
ガタン・ガンガン・こつこつ・びいびい  
ガタガタ・ガサガサ・ボンボン・パタパタ・ビリビリ

★判別したい文を入力してください★  
ガンガン勉強する

送信 リセット



2:21  
ark4.media.eng.hokudai.ac.jp

Thank you for using :)

あなたが入力した文  
「ガンガン勉強する」  
に存在するオノマトペ  
【ガンガン】  
の意味は  
<<頭が痛い様子、たくさん~する様子などがあります。擬態語です。>>  
同じように使われる例として  
★それでも、本番はいつも通りガンガンとばせるように、心がけていますが。  
といった例文が挙げられます

もう一回やる



# ルールベースシステムの精度

(全1350文による実験)

	正解率( $\gamma / \alpha$ )	適合率( $\gamma / \beta$ )	F値
ガサガサ	72.3	83.9	77.6
ガタガタ	60.7	79.2	68.8
パイパイ	65.3	75.0	69.8
ビリビリ	73.8	84.0	78.6
ガタン	73.0	81.8	77.1
ポンポン	57.4	69.7	63.0
コツコツ	71.2	90.4	79.7
ガンガン	66.2	79.3	72.2
バタバタ	66.2	90.7	76.6
合計	67.3	81.4	73.7

$\alpha$  = 人手により正解が付与できた文の総数

$\beta$  = ルールベースシステムにより答えが付与された文の総数

$\gamma$  = システムが与えた出力が正解だった文の総数



## 解析エラー

名詞が抽出できなかった場合も含め

正解を出力したい



意味決定単語が抽出できない場合は

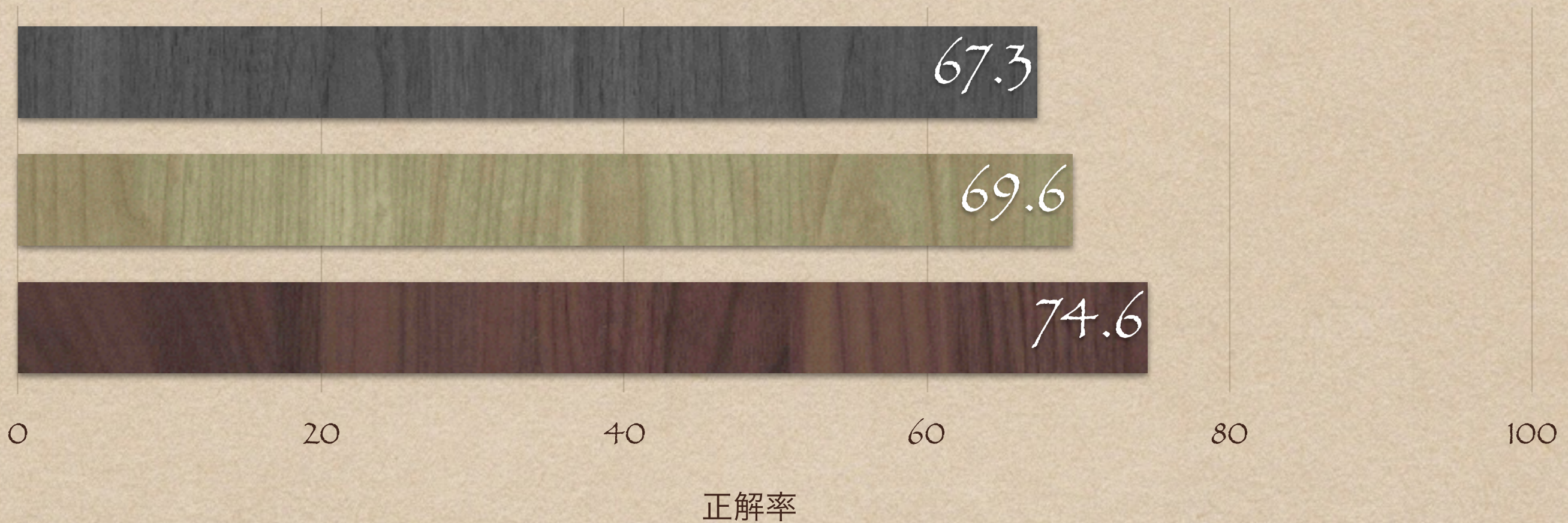
機械学習を使用する

(人が与えた正解データを元に)



# 提案システムの性能（正解率）

■ ルールのみ      ■ 機械学習のみ      ■ 機械学習＋ルール



機械学習とルールを取り入れることで  
判別システムとして一定の精度を得た




# まとめと今後の課題

- ・ オノマトペの意味判別に**名詞の性質**を用いることは有効.
- ・ **ルールと機械学習**を組み合わせ、オノマトペの意味判別を、**高精度**で可能とした.
- ・ エラーは**係り受け解析の失敗**、**カテゴリを与えられない新規名詞**などが存在が挙げられた.  
→解決し実用化(日本語教育など)へ向け、さらに高精度へ



北海道大学  
HOKKAIDO UNIVERSITY



Thank you for your attention!!