

## Introduction



機械の  
言語理解  
口語の  
意味理解

同じオノマトペが  
全く異なる意味で使用される  
場合があるため、  
文中でどの意味で使われているか  
判別する必要がある！

機械翻訳



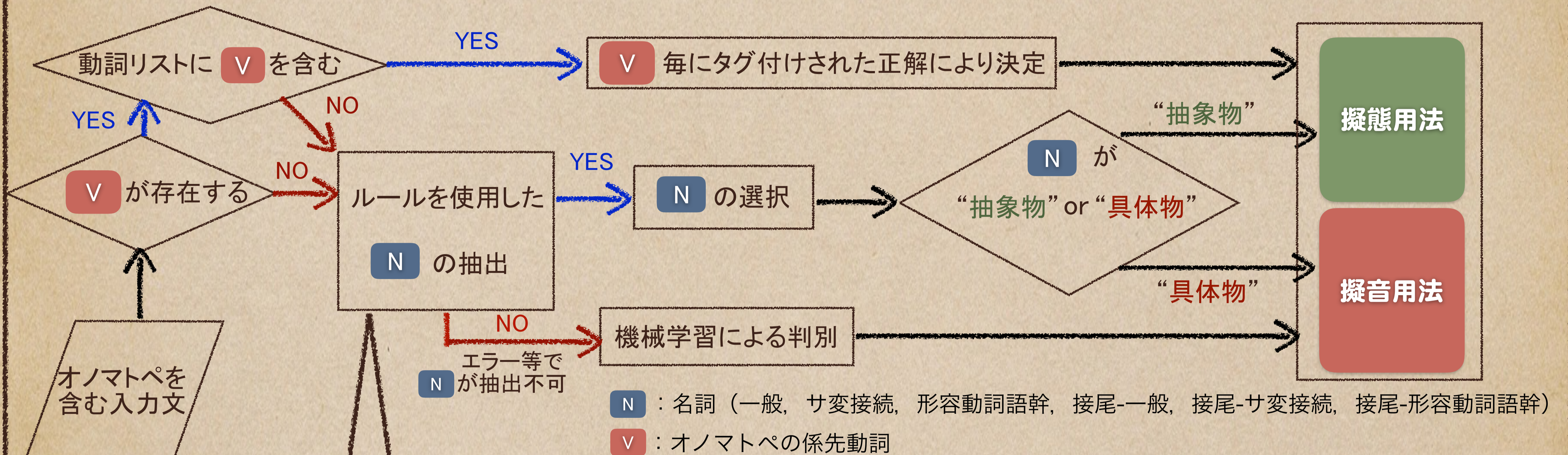
急激に



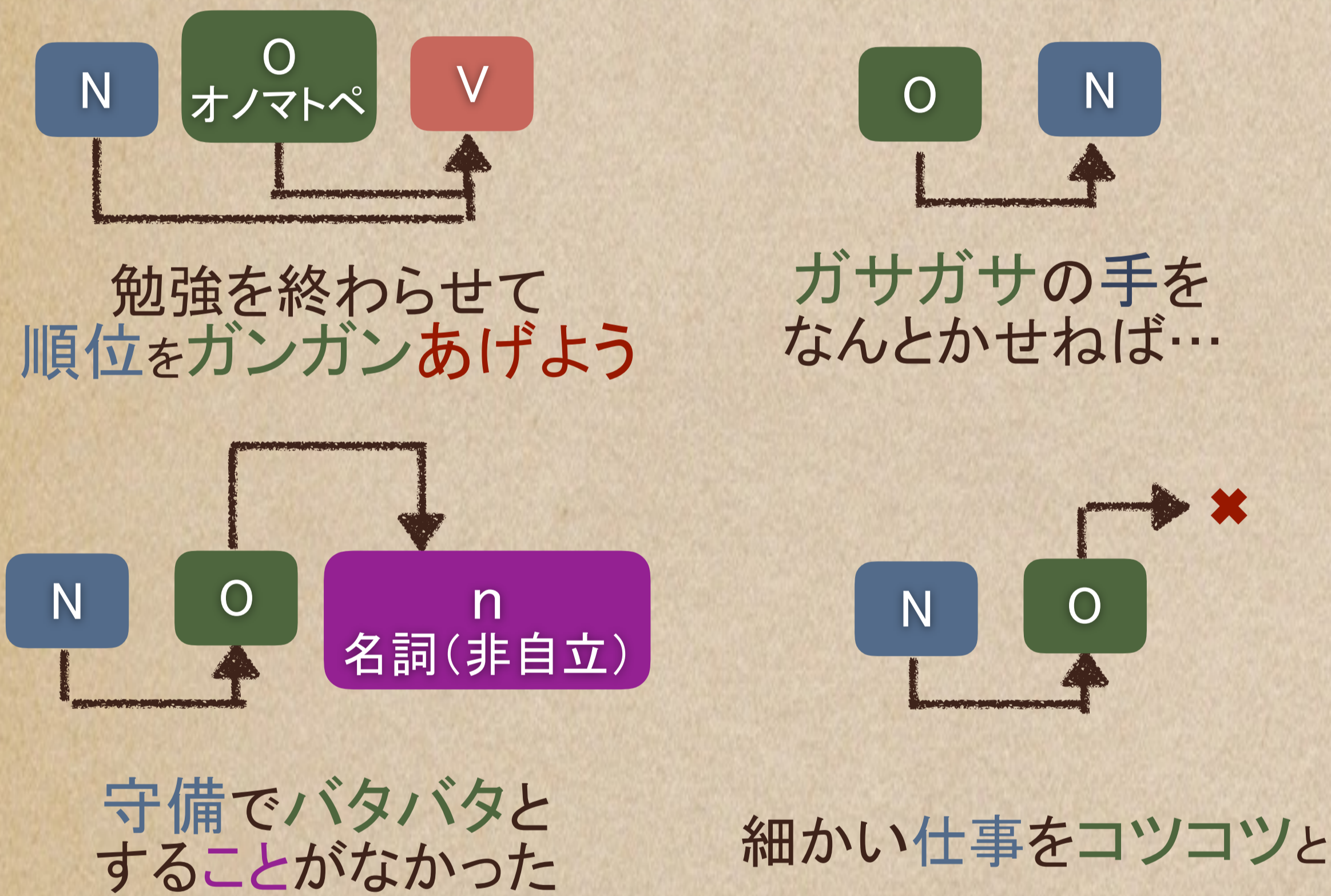
体調がガタンと落ちた

## Method

※動詞リスト: V とオノマトペの共起のみで意味が判別できるような V の集合

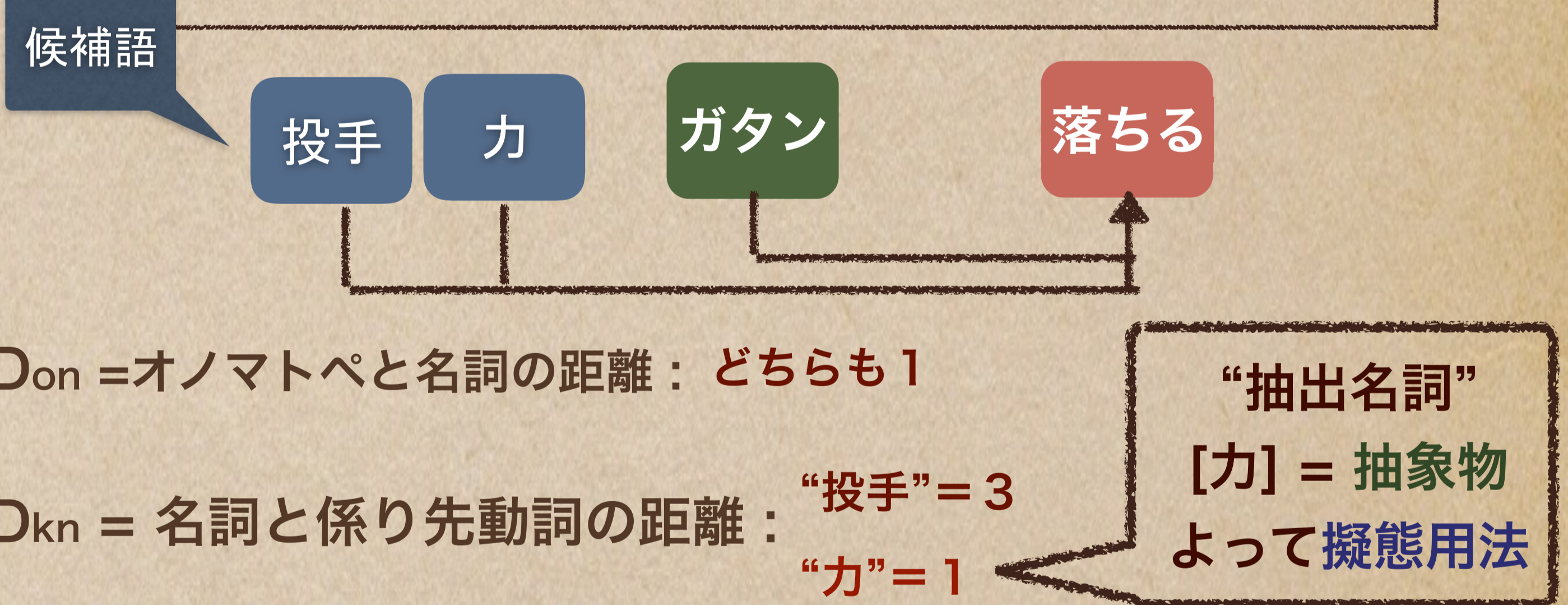


## What rules?



## For example...

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

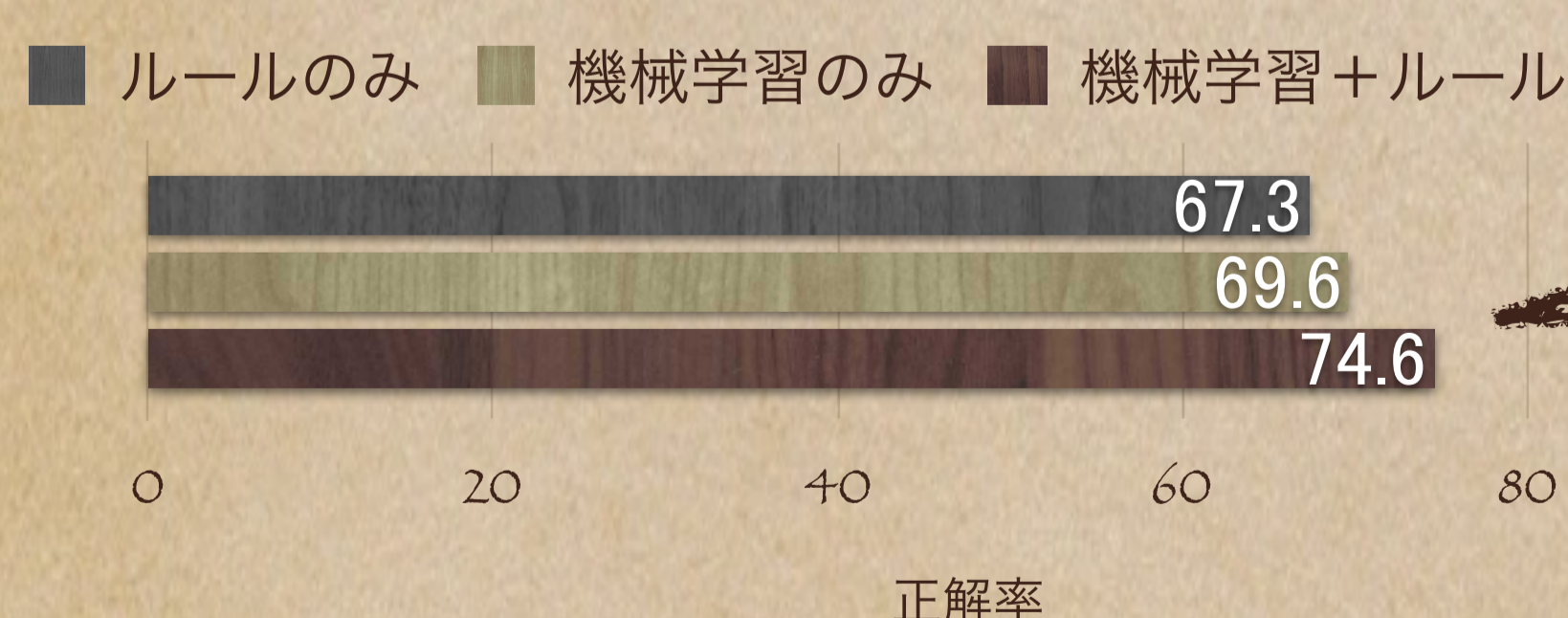


## Results

★ルールのみを適用したシステムの適合率( $\gamma/\beta$ )は81.4%であり、  
名詞を用いたルールの有効性が示された。

★文の収集元はTwitter<sup>1</sup>・アメーバコーパス(オノマトペが多数出現する文は口語文の割合が高いため)  
としたが、口語調であるが故に<カテゴリが付与されず分類できない名詞の出現>や  
<係り受け解析のエラーにより意味判別名詞 N が抽出できない>等、意味判別の  
失敗が見られた。

エラー文を機械学習で判別する  
システム(機械学習+ルール)を提案



・ルールのみより7.3pt上昇  
・機械学習とルールを組み合わせた「提案システム」が最も高い正解率

システム別正解率( $\gamma/\alpha$ )の比較 (全1350文による実験)

| 単位[%] | ルールのみ | ルール+機械学習 |
|-------|-------|----------|
| ガサガサ  | 72.3  | 82.5     |
| ガタガタ  | 60.7  | 78.6     |
| パイパイ  | 65.3  | 69.4     |
| ビリビリ  | 73.8  | 82.6     |
| ガタン   | 73.0  | 80.4     |
| ポンポン  | 57.4  | 66.9     |
| コツコツ  | 71.2  | 71.2     |
| ガンガン  | 66.2  | 70.3     |
| パタパタ  | 66.2  | 70.3     |
| 合計    | 67.3  | 74.6     |

$\alpha$  = 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

## Summary

- ・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だと証明した。
- ・名詞の性質を用いて作成したルール、先行研究<sup>2</sup>より有効だと証明された名詞や係り先動詞を素性とした機械学習を組み合わせ、75%程度の高い正解率をもつ意味判別システムを構築した。
- ・今後意味判別システムを、言語理解や言語教育に実用化するため、ルールの追加等を行い、さらに精度を高めたい。

1) Twitter : <https://twitter.com/>

2) 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