

# Webデータに基づく複合動詞データベースの構築

山口昌也  
(国立国語研究所)

## 概要

### 構築の背景

#### 複合動詞と構成動詞の関係分析

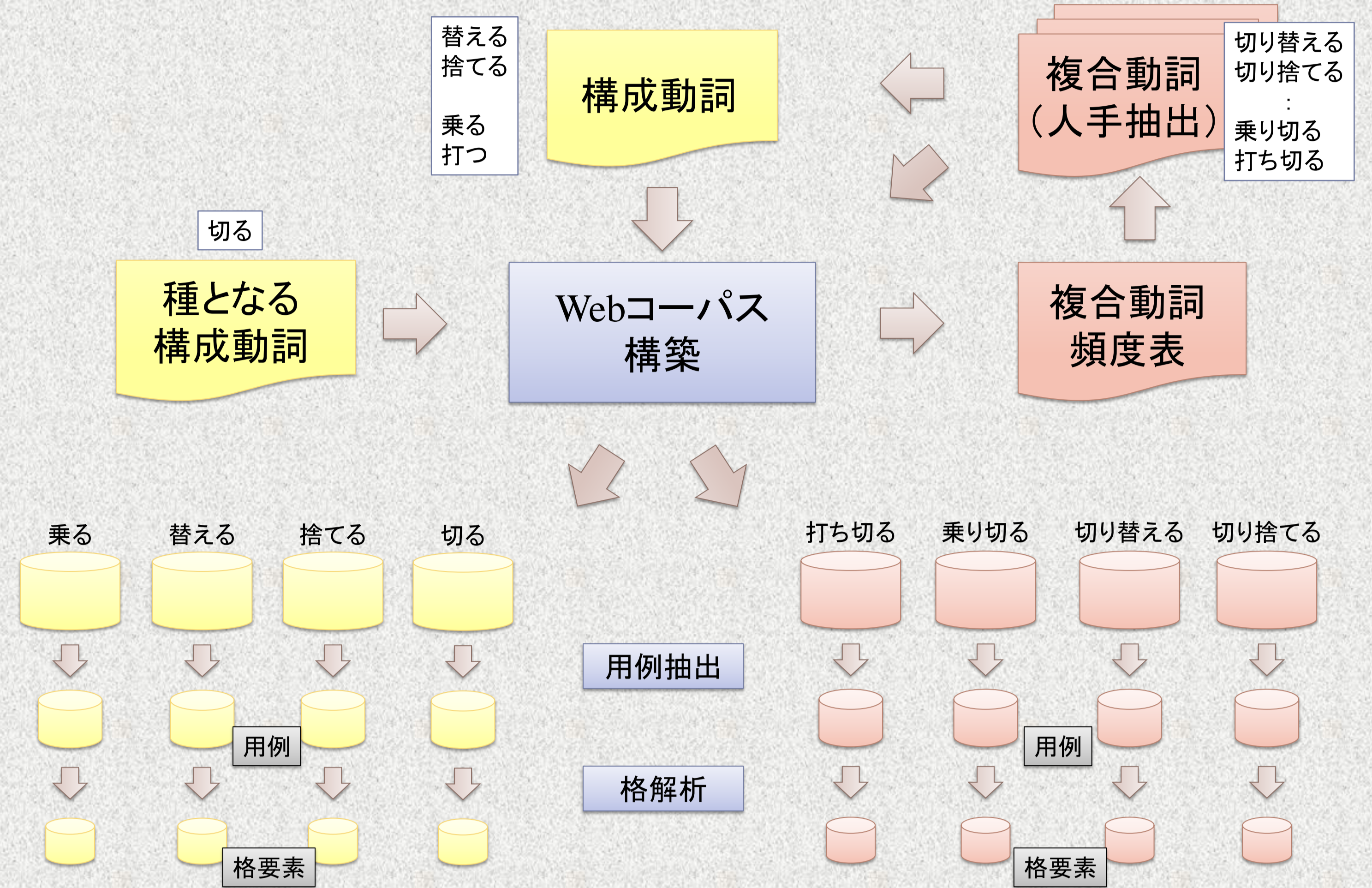
投げる		すべてが継承されるわけではない	投げ込む	
ヲ格	ニ格		ヲ格	ニ格
ボール	相手	ボール	中	
球	どこ	球	川	
石	遠く	速球	海	
● 疑問	実際	直球	池	
● 身	時	石	そこ	
物	中	ストレート	口	
ルアー	海	手榴弾	山	
● 質問	上	スライダー	水面	
● 言葉	人	瓶	ポスト	
速球	ところ	ルアー	客席	

すべてが構成動詞に由来するのかわ?

### 構築の目的

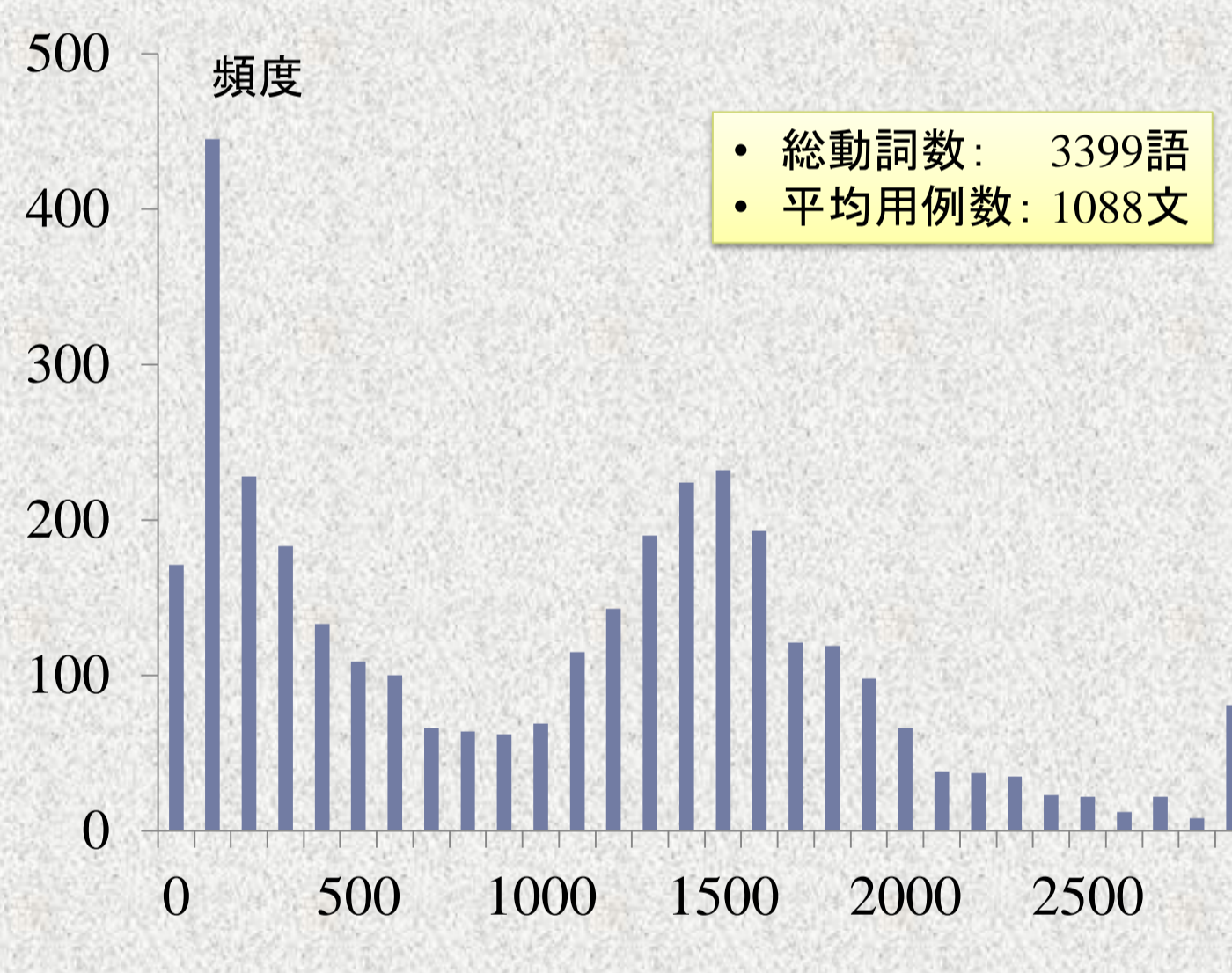
- 実例に基づいた、複合動詞・構成動詞間の関係分析をすること
  - ⇒ 一定数以上の用例を収集可能な複合動詞を収録する
  - ⇒ Webデータの特徴に合わせて、用例を収集・格解析する

## 構築方法

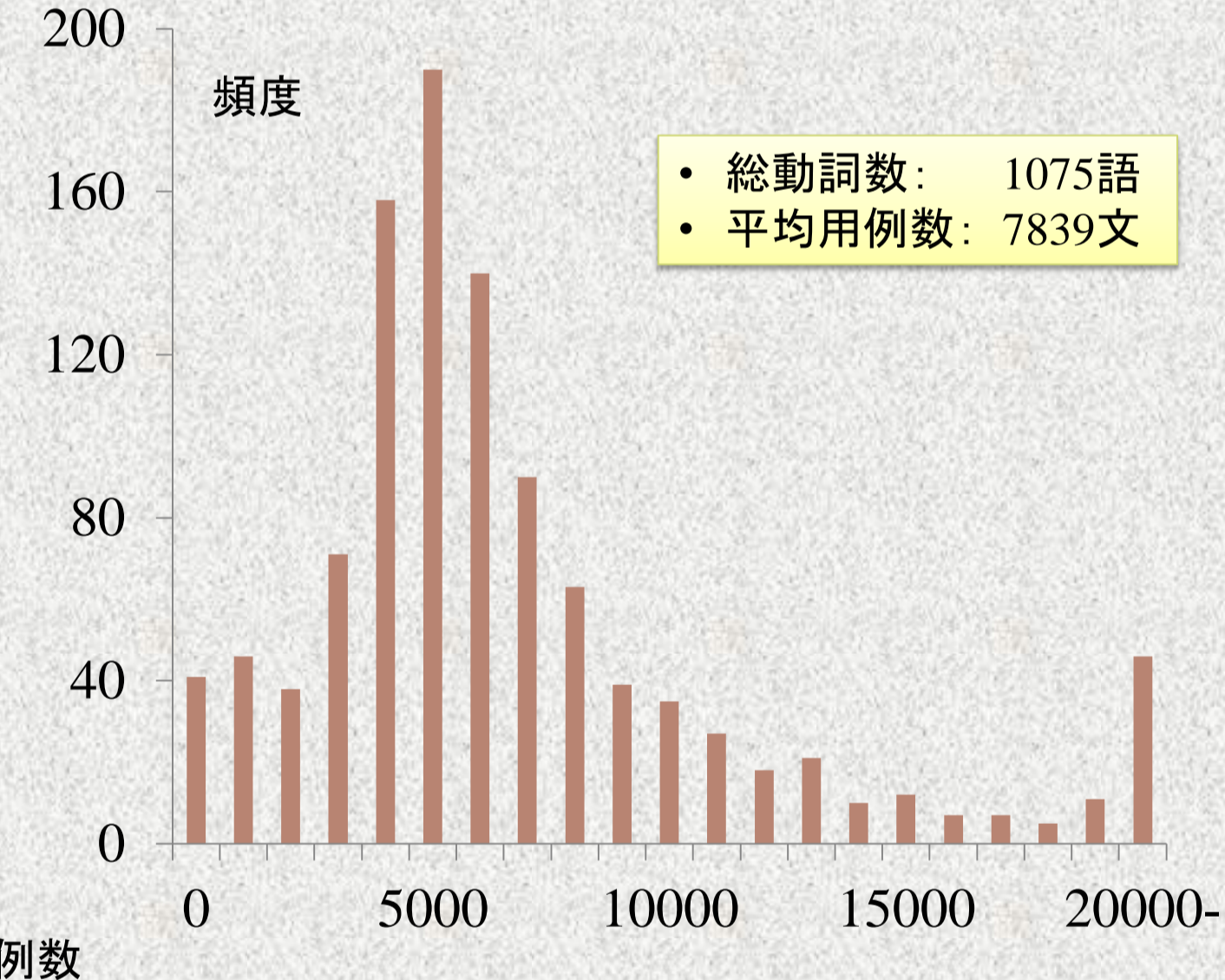


## 構築結果 (http://csd.ninjal.ac.jp/comp)

### 収集結果(複合動詞)



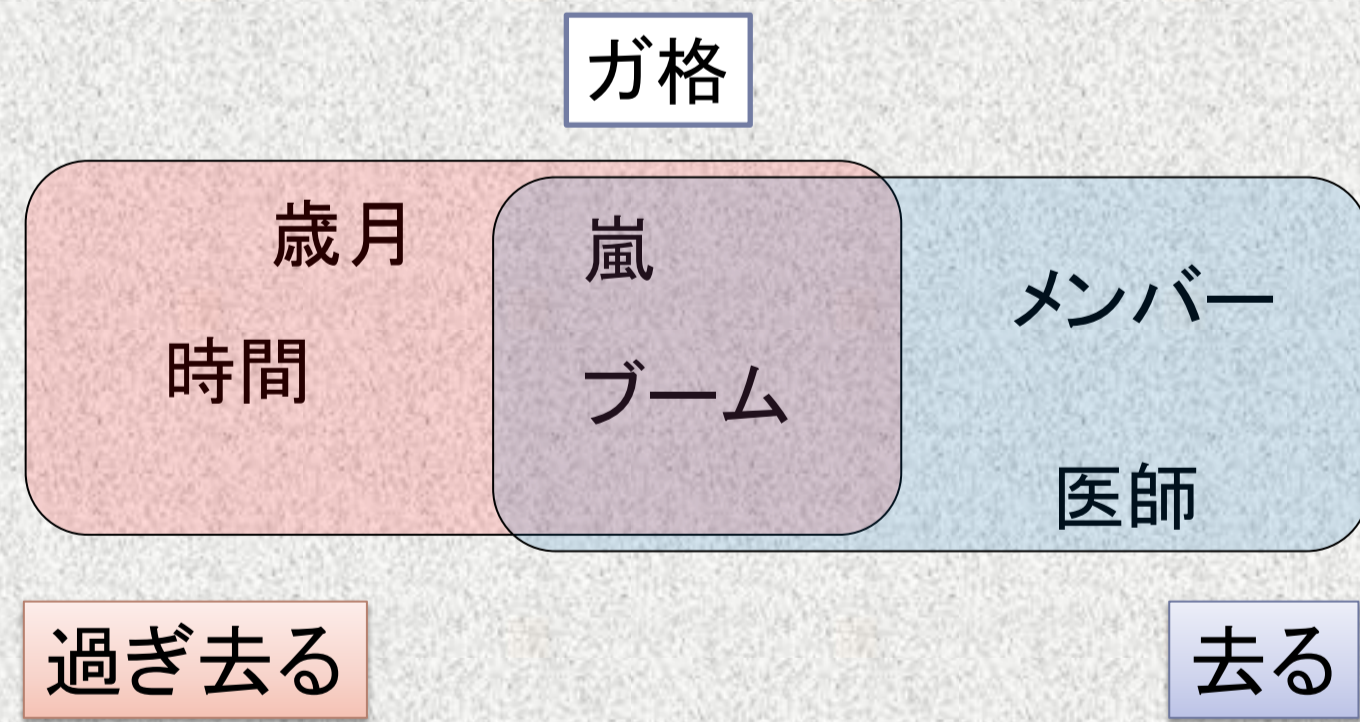
### 収集結果(構成動詞)



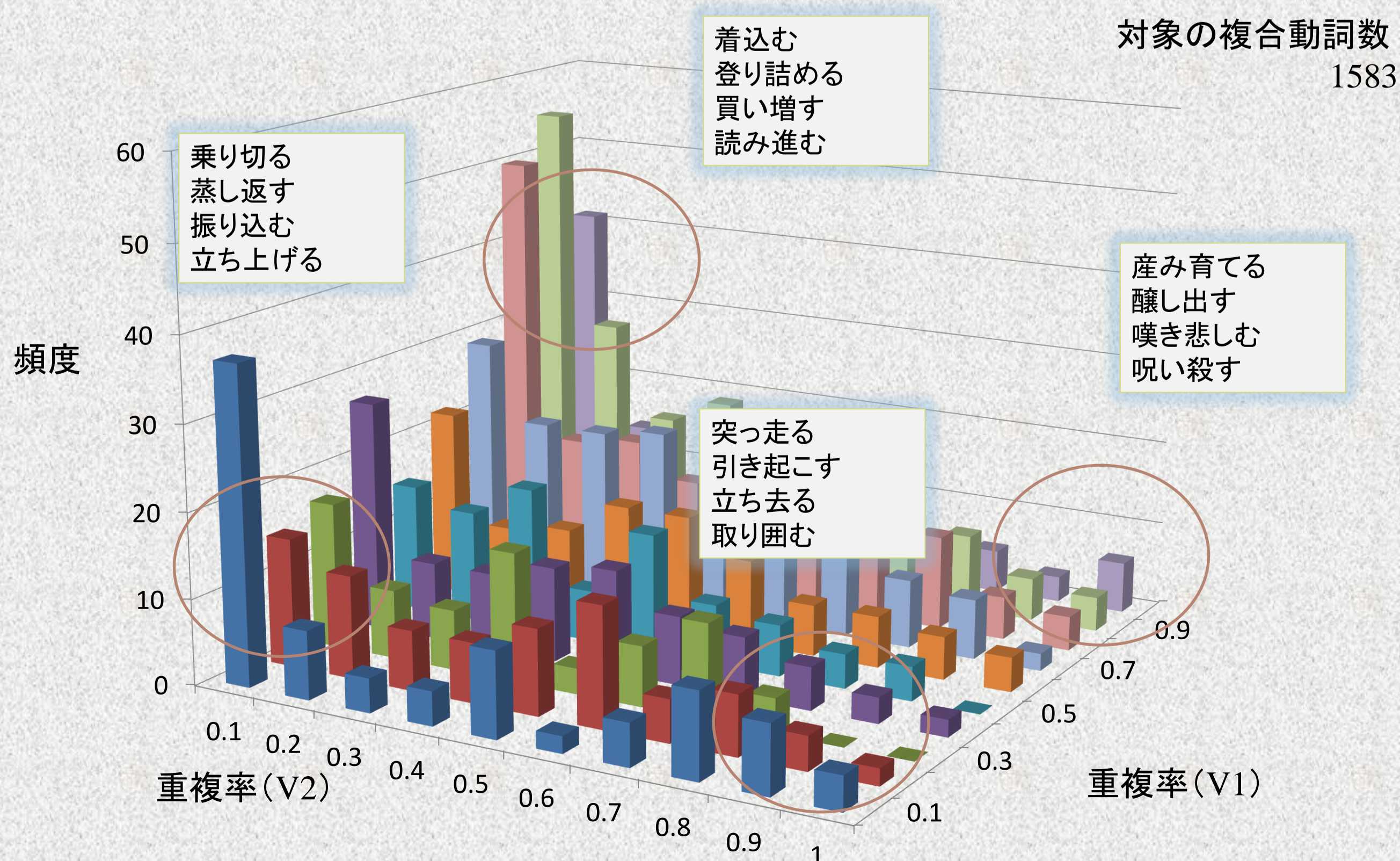
### 重複度

- 複合動詞の格iの取りうる格要素が、構成要素の格iの格要素となりうる割合

$$OV_i = \sum_{w_a \in E_{ct} \cap E_{si}} n(w_a) / \sum_{w_b \in E_{ct}} n(w_b)$$



### 重複度の分布(ヲ格, F<sub>複,ヲ格</sub> ≥ 50, F<sub>複</sub> ≥ 1000, F<sub>単</sub> ≥ 2000)

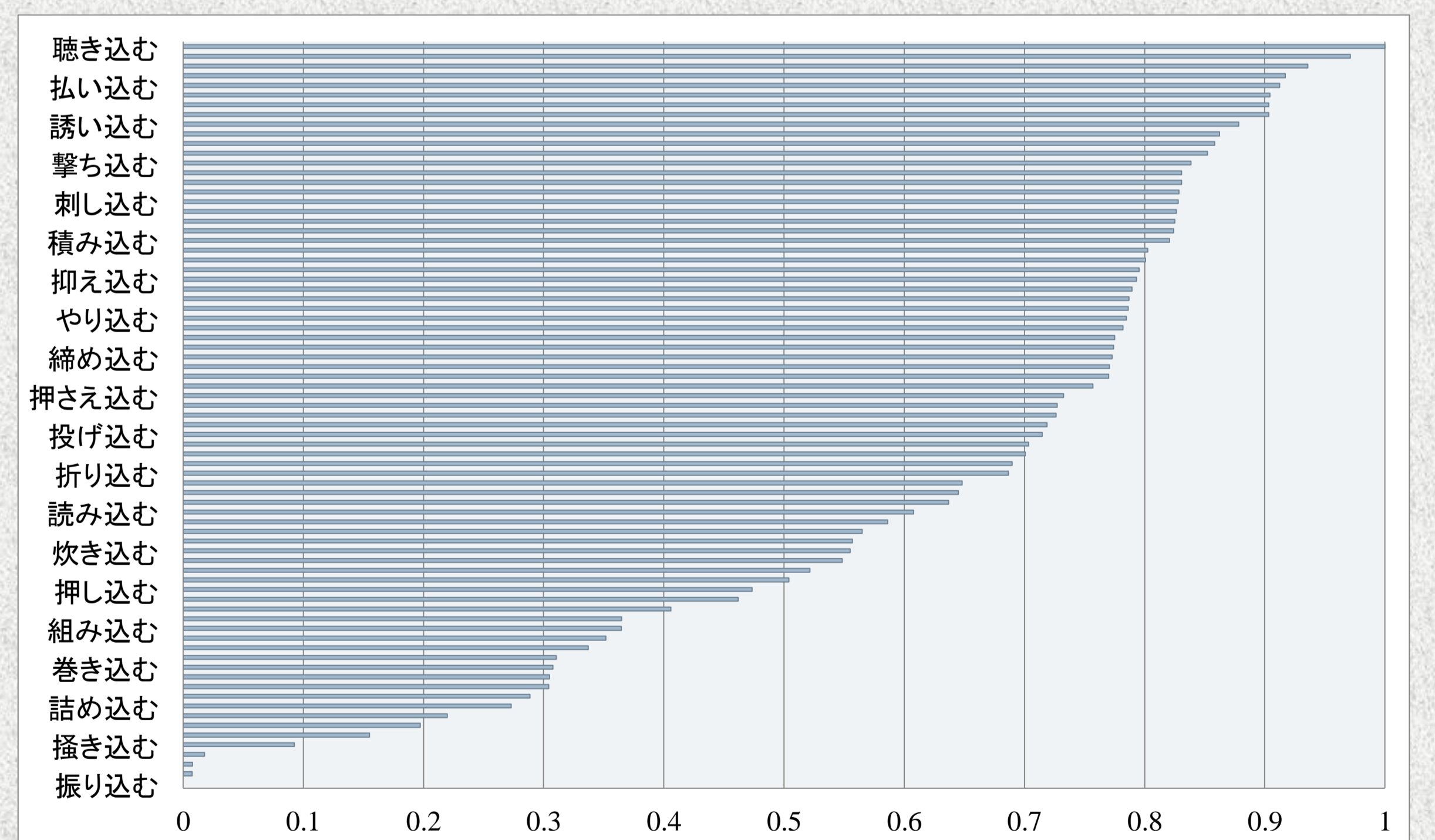


## 分析例: 「～込む」と前項動詞との関係分析

### 分析の狙い

- 先行研究 ⇒ 複合動詞となったときの変化の記述が不十分
  - 格支配構造の関係の有無(山本1984)
  - 「払い込む」(前項のみ関係あり), 「振り込む」(前後とも関係なし)
  - LCS(語彙概念構造)の合成(由本2005)
  - 「嘆き悲しむ」LCS1 and LCS2
  - 格要素の変化を分析

### 重複度の測定結果(ヲ格, F<sub>複,ヲ格</sub> ≥ 50, F<sub>複</sub> ≥ 1000, 99語)



### 関係分類

- 継承
    - 「聞き込む」(1.0)「着込む」(0.97)
    - 「擦り込む」(0.94)
  - 別義
    - 「振り込む」(0.0)「擦り込む」(0.01)
    - 「申し込む」(0.01)
  - 派生
    - 別義が混在
    - 「織り込む」(OV<sub>ヲ</sub> = 0.29)
    - 「糸を織り込む」 ⇔ 「糸を織る」
    - 「情報を資料に織り込む」 ⇔ \*「情報を織る」
- 分布変化  
構成動詞では、あまり用いない格要素  
⇒ 変化大
- 「流し込む」(OV<sub>ヲ</sub> = 0.46)  
「鉄を鋳型に流し込む」 ⇔ ?「鉄を流す」
- | 格要素  | 複合動詞 | 構成動詞 |
|------|------|------|
| モルタル | 19   | 0    |
| 樹脂   | 17   | 0    |
| ビール  | 17   | 0    |
| 金属   | 16   | 0    |
| セメント | 15   | 0    |
| コンクリ | 13   | 0    |
| 鉄    | 11   | 0    |