

# 動詞連想概念辞書の構築とその特徴

寺岡文博(慶應義塾大) 岡本潤(SFC研究所) 石崎俊(慶應義塾大)

## 研究の目的

- 言葉の背景にある膨大な知識を体系化
  - 動詞を中心
  - 人間の直感に基づいた連想実験データ
- コンピュータによる言語意味理解の高精度化
  - 人間の連想に基づいた言語処理に応用
  - 連想概念辞書の応用

## 連想実験

- 基本動詞798語 → 刺激語54語
  - 小学校の国語の教科書
  - 他動詞の優先など
- 連想課題
  - 動作主 対象 始点 終点 時点 場所 手段 様相 理由 目的
- 1刺激語1連想語につき被験者40人

## 連想距離の定量化

- 線形計画法による最適化

$$D = \frac{7}{10} \times F + \frac{1}{3} \times S + 0 \times T$$

$$F = \frac{N}{n + \delta} \quad n = \text{連想人数}, n \geq 1$$

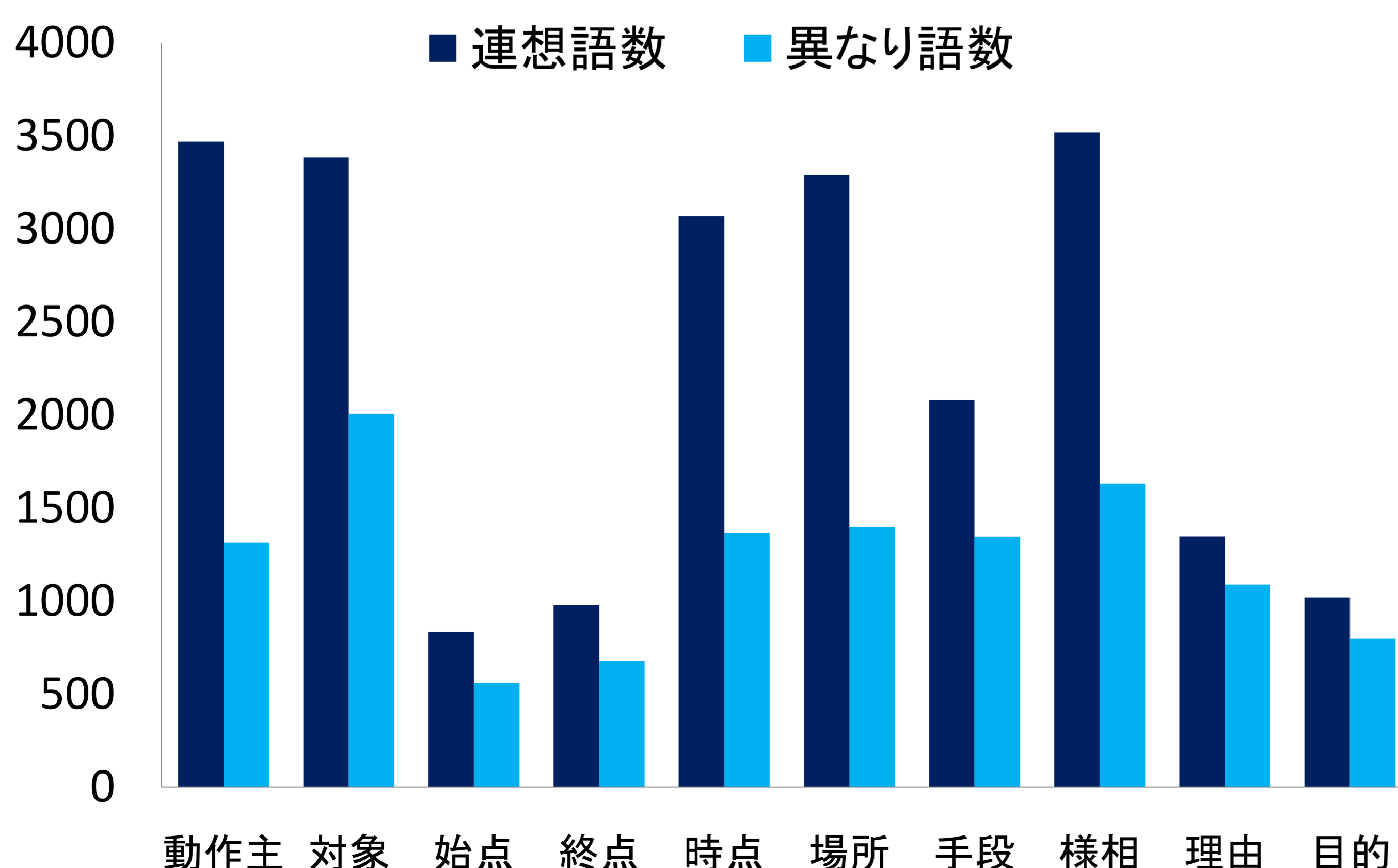
$$\delta = \frac{N}{10} - 1 \quad (N \geq 10) \quad N = \text{被験者数}$$

$$S = \frac{1}{n} \sum_{i=1}^n s_i \quad s_i = \text{被験者が連想した語の順位}$$

$$T = \log_{10} \left( \frac{1}{n} \sum_{i=1}^n t_i \right) \quad t_i = \text{被験者が連想に要した時間}$$

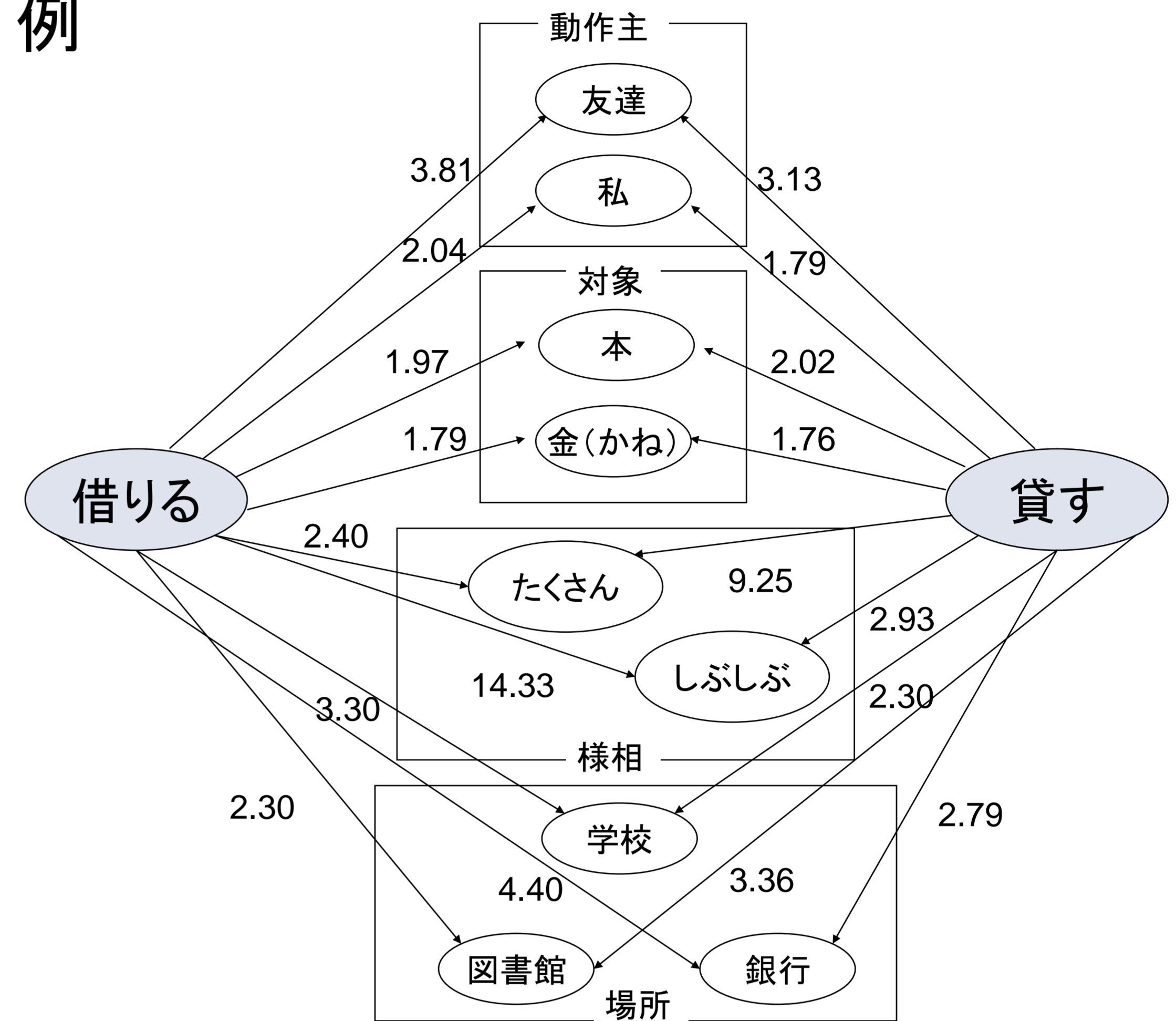
## 動詞連想概念辞書

- 連想語と異なり語



## 連想語と連想距離

- 例



## 大規模格フレームとの比較

- 大規模格フレームβバージョン
  - Web上の日本語テキスト(約5億文)
  - 約5万用言に対する格フレームと格要素
  - 自動的に構築

<http://nlp.kuee.kyoto-u.ac.jp/nl-resource/casframe.html>

- スピアマンの順位相関係数( $r_s$ )
  - 共通する上位10語
  - 「動作主」と「ガ格」「対象」と「ヲ格」

動詞	$r_s$	動詞	$r_s$
貸す	0.687879	傷める	0.936364
建てる	0.636364	耕す	0.818182
走る	0.090909	建てる	0.700000
買う	0.090909	借りる	0.169697
滑る	-0.457580	貸す	-0.00303

## 考察

- 連想課題「動作主」と格フレーム「ガ格」
  - 特有の動作主と「私」「人」
- 連想課題「対象」と格フレーム「ヲ格」
  - 対象の属性と被験者の偏り
- 動詞まわりの意味情報の表出
  - 小規模なデータ
  - 連想距離の妥当性

## 今後の課題

- 動詞連想概念辞書データの拡充
- 連想語の表記修正
- 比較対象の広範囲化