



大規模コーパスに基づく言語分析

～日本語二重目的語構文の基本語順の分析を例に～

笹野遼平（東工大）

自己紹介

略歴

- 東大/京大 黒橋研 (2003~2010)
- 東工大 奥村・高村研 助教(2010~)

研究の興味

- 自然言語処理の基礎解析
 - 形態素解析における未知語処理(JUMAN6.0~)
 - 固有表現認識・ゼロ照応解析・連想照応解析(KNP4.0~)
- 言語現象の分析
 - 格交替に関する語彙知識の自動獲得(EMNLP'13)
 - 二重目的語構文の基本語順の分析(ACL'16)

```

> echo "こんにちわ*面白い占いサイト作ってまあーすう*楽しくないかもだけえどお 楽しんでね*" | juman
こんにちわ こんにちわ こんにちわ 名詞 6 普通名詞 1 * 0 * 0 "自動獲得:テキスト 非標準表記"
☆☆ 特殊 1 記号 5 * 0 * 0 NIL
面白い おもしろい 面白い 形容詞 3 * 0 イ形容詞アウオ段 18 基本形 2 "代表表記:面白い/おもしろい 反義:形容詞:つまらない/つまらない"
占い うらない 占う 動詞 2 * 0 子音動詞口行 12 基本連用形 8 "代表表記:占う/うらなう"
サイト さいと サイト 名詞 6 普通名詞 1 "代表表記:サイト/さいと カテゴリ:場所-その他 ドメイン:家庭・暮らし"
作って つくって 作る 動詞 2 * 0 基本連用形 14 "代表表記:作る/つくる"
まあーすう ます ます 接尾辞 1 "接尾辞ます型 31 基本形 2 "代表表記:ます/ます 長音挿入"
☆☆ 特殊 1 記号 5 * 0 * 0 NIL
楽しく たのしく 楽しい 形容詞 3 * 0 イ形容詞イ段 19 基本連用形 7 "代表表記:楽しい/たのしい"
ない ない ない 接尾辞 14 形容詞 1 "代表表記:ない/ない 非標準表記"
か か か 助詞 9 接続助詞 3 * 0
も も も 助詞 9 副助詞 2 * 0
だ だ だ 判定詞 4 * 0 判定詞 25 基本形 2 NIL
けえどお けど けど 助詞 9 接続助詞 3 * 0 * 0 "長音挿入"
☆☆ 特殊 1 空白 6 * 0 * 0 NIL
楽しんで たのしんで 楽しむ 動詞 2 * 0 子音動詞マ行 9 タ系連用形 14 "代表表記:楽しむ/たのしむ 形容詞派生:楽しい/たのしい"
ね ね ね 助詞 9 終助詞 4 * 0 * 0 NIL
☆☆ 特殊 1 記号 5 * 0 * 0 NIL
EOS

```

小書き文字化

長音挿入

研究の興味

– 自然言語処理の基礎解析

- 形態素解析における未知語処理 (JUMAN6.0~)
- 固有表現認識・ゼロ照応解析・連想照応解析 (KNP4.0~)

– 言語現象の分析

- 格交替に関する語彙知識の自動獲得 (EMNLP'13)
- 二重目的語構文の基本語順の分析 (ACL'16)

```

> echo "昨日、太郎はケーキを買って食べていた。\\nちょっと価格は高そうだったけれど、とてもおいしそうだった。" | juman | knp -anaphora
# S-ID:1 KNP:4.17-CF1.1 DATE:2016/05/16 SCORE:-22.23185
  昨日n、          <体言><NE:DATE:昨日><EID:0>
  太郎Jはp        <体言><NE:PERSON:太郎><EID:1>
  ケーキnをp      <体言><EID:2><述語項構造:ケーキ/ケーキ:名1>
  買ってv         <用言:動><EID:3><述語項構造:買う/かう:動2:ヲ/C/ケーキ/2;ガ/O/太郎/1>
  食べてvいたs。 <用言:動><EID:4><述語項構造:食べる/たべる:動1:時間/C/昨日/0;ガ/N/太郎/1;ヲ/O/ケーキ/2>
EOS
# S-ID:2 KNP:4.17-CF1.1 DATE:2016/05/16 SCORE:-32.30108
  ちょっとa      <EID:5>
  価格nはp       <体言><EID:6><述語項構造:価格/かかく:名1:ノ/O/ケーキ/2>
  高jそうだったsけれどp、 <用言:形><EID:7><述語項構造:高い/たかい:形5:修飾/C/ちょっと/5;ガ/N/価格/6>
  とてもa       <EID:8>
  おいしjそうだったs。 <用言:形><EID:9><述語項構造:美味しい/おいしい:形1:修飾/C/とても/8;ガ/O/ケーキ/2>
EOS

```

固有表現認識

連想照応解析

ゼロ照応解析

研究の興味

- 自然言語処理の基礎解析
 - 形態素解析における未知語処理 (JUMAN6.0~)
 - 固有表現認識・ゼロ照応解析・連想照応解析 (KNP4.0~)

- 言語現象の分析
 - 格交替に関する語彙知識の自動獲得 (EMNLP'13)
 - 二重目的語構文の基本語順の分析 (ACL'16)

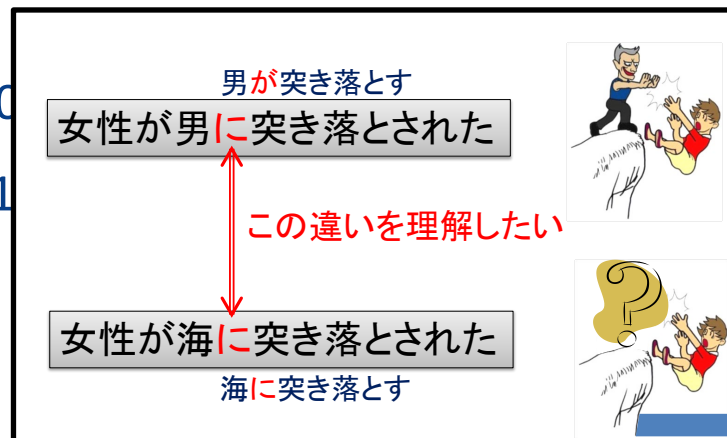
自己紹介

略歴

- 東大/京大 黒橋研 (2003~2010)
- 東工大 奥村・高村研 助教 (2011~)

研究の興味

- 自然言語処理の基礎解析
 - 形態素解析における未知語処理 (JNLP 6.0)
 - 固有表現認識・ゼロ照応解析・連文照応解析
- 言語現象の分析
 - 格交替に関する語彙知識の自動獲得 (EMNLP'13)
 - 二重目的語構文の基本語順の分析 (ACL'16)



言語学の知見 +
大規模コーパスから
得られる語彙知識

今日の話

- 言語現象に対する各種アプローチ (≈ 30%)
 - 言語学からのアプローチ
 - 脳科学・認知科学からのアプローチ
 - 自然言語処理(NLP)/計算言語学(CL)からのアプローチ
- 日本語二重目的語構文の基本語順の分析 (≈ 70%)
 - 日本語二重目的語構文と既存のアプローチ
 - 大規模コーパスに基づく基本語順の分析

言語現象に対するアプローチ

(理論)言語学

- 現象を説明できる理論を構築
- ➡ 理論を支持する例を基に検証, etc.

言語処理/計算言語学

1. 言語現象をモデル化
➡ 精度により検証 **or**
2. 仮説+自動用例収集
➡ 大規模コーパスに基づく検証, etc.

興味のある
言語現象

脳科学

- 脳の反応に関する仮説を立てる
- ➡ 基本状態との有意差を実験的に検証, etc.

マルチモーダルなアプローチ

- [Ordonez+, ICCV'13] From Large Scale **Image** Categorization to **Entry- Level Categories**
- [Shutova+, NAACL'16] Black Holes and White Rabbits: **Metaphor Identification** with **Visual** Features

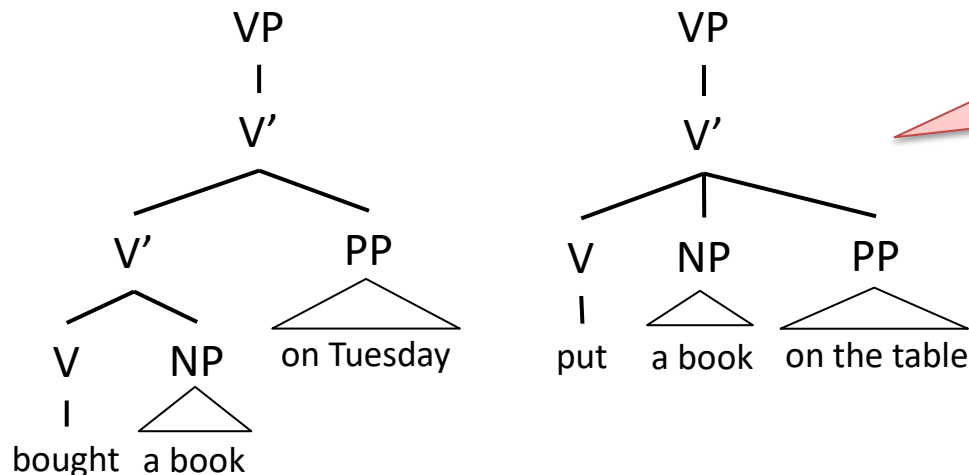
言語学的アプローチ(個人的印象)

😊 既存理論で説明できない用例の発見

😊 該当用例を上手く説明できる理論の構築

e.g. 項と付加詞の区別が必要となる根拠 (渡辺明「生成文法」3.2節より)

- John bought a book on Tuesday, but Sam did so on Friday.
- *John put a book on the table, but Sam did so on the desk.



“did so”で置き換えることができるのはV'の部分のみ

言語学的アプローチ(個人的印象)

😊 既存理論で説明できない用例の発見

😊 該当用例を上手く説明できる理論の構築

e.g. 項と付加詞の区別が必要となる根拠 (渡辺明「生成文法」3.2節より)

- John bought a book **on Tuesday**, but Sam **did so** on Friday.
- *John put a book **on the table**, but Sam **did so** on the desk.

e.g. 間接・直接受動の二格の統語的性格 (渡辺明「生成文法」4.4節より)

- 太郎が**花子**に**自分**の部屋で泣かれた [自分=太郎,花子]
- 太郎が**花子**に**自分**の部屋で叱られた [自分=太郎]

😞 多くは1人~少人数の容認性判定に依存 (cf. corpus-based)

😞 多数の組み合わせを考える必要がある現象に不向き

脳科学からのアプローチ(個人的印象)

😊 もっとも直接的な分析手法

😊 個人の直感・内省に依存しない検証が可能

主な手法: fMRI(機能的磁気共鳴映像法), MEG(脳磁計測), etc.

- [Chang+, ACL'09] Quantitative modeling of the neural representation of adjective-noun phrases to account for **fMRI** activation
- [Fyshe+, ACL'14] Interpretable Semantic Vectors from a Joint Model of **Brain-** and Text-Based Meaning

😞 実験のコストが非常に大きい & コントロールが難しい

😞 有意差が出た時点で打ち切ることが多い

😞 多数の組み合わせを考える必要がある現象に不向き

大規模コーパスに基づくアプローチ

- 😊 極めて大規模なデータに基づく網羅的な検証が可能
- 😊 多数の組み合わせを考慮した分析が低コストで可能
 - 格交替に関する知識獲得、二重目的語構文の語順分析
- 😊 少数の人の直感・内省に依存しない
 - 実際に使用されている用法＝“正しい”言葉
- 😞 各データの信頼性は高くない
 - 実際に観測される表現は多くの要因が作用した結果
 - 多くの解析誤りや想定していないバイアスの存在
- 😞 理論的分析は後回しとなる傾向

各種アプローチの比較

	コスト	大規模化	客観性	信頼性	直接性
NLP/CL	😊	😊	😊	😞	😞
言語学	😐	😞	😞	😊	😐
脳科学	😞	😞	😊	😐	😊



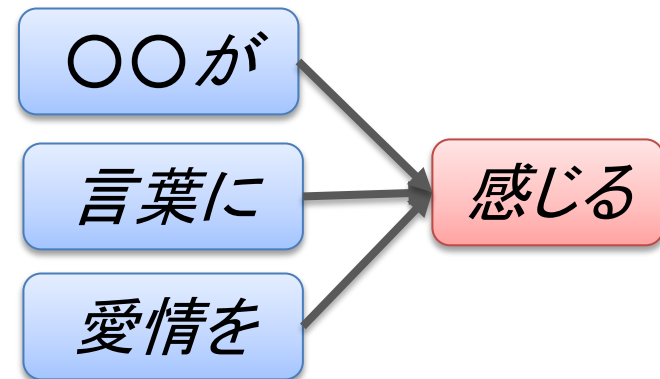
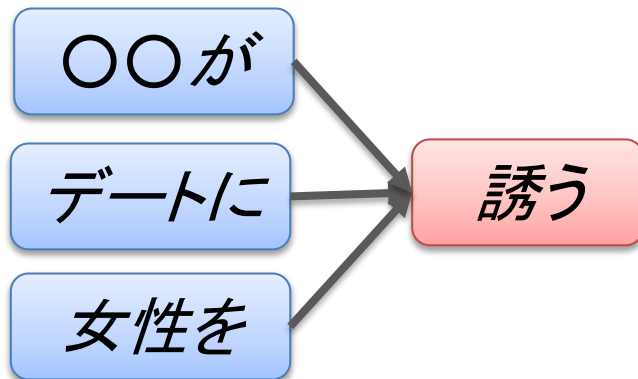
- 多数の組み合わせを考慮すべき現象の分析:
 1. 大規模コーパスに基づく分析で仮説を立てる (by NLP/CL)
 2. 言語学的分析・脳科学的計測で仮説を検証という手順が良いのでは？

今日の話

- 言語現象に対する各種アプローチ
 - 言語学からのアプローチ
 - 脳科学・認知科学からのアプローチ
 - 自然言語処理(NLP)/計算言語学(CL)からのアプローチ
- **日本語二重目的語構文の基本語順の分析**
 - 日本語二重目的語構文と既存のアプローチ
 - 大規模コーパスに基づく基本語順の分析

二重目的語構文

- 項として2つの目的語を取る構文
 - 日本語の場合、各目的語は**ヲ格**と**ニ格**で表される
 - **ガ格**も含めると項は3つ ⇒ **三項動詞文**とも呼ばれる



二重目的語構文の基本語順

• 「～に～を」なのか？

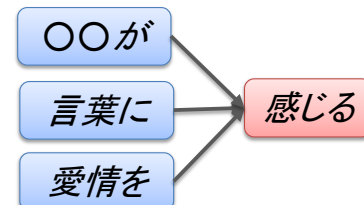
コーパス中の
出現割合

言葉に愛情を感じる

(97.5%)

タイヤに空気を入れる

(99.1%)



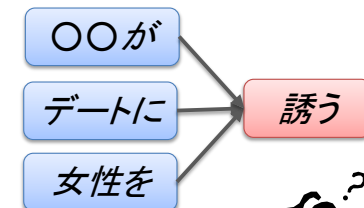
• 「～を～に」なのか？

女性をデートに誘う

(99.6%)

子供を風呂に入れる

(98.6%)



⇒ 基本語順を決める要因を明らかにしたい

本研究で前提とする立場

- 基本語順は語彙により異なるという前提
 - 多くの先行研究で採用 [Matsuoka'03, Miyagawa+'04, 中本+'06, etc.]
 - ただし、二重目的語構文の基本語順は1つという立場も [Hoji'85]
 - 動詞と2つの目的語が与えられた場合に、ある語順が大半を占めるならば、その語順が基本語順であるとみなす 女性をデートに誘う
(99.6%)
- 工学的な有用性よりも言葉への興味
 - 人は何を根拠に特定の語順を自然と感じるのか？
 - 意味役割の自動認識や文生成の手掛かりとなる可能性は考えられる
(ただし、実際の文生成では、項の長さや情報構造が重要)

二重目的語構文の基本語順に関する先行研究

- 言語学や脳科学からのアプローチ
 - 言語学者/被験者の内省に基づく分析
e.g. [Hoji'85, Matsuoka'03, Miyagawa+'04] / [中本+'06]
 - 読み時間/脳波等の計測に基づく分析
e.g. [Koizumi+'04, Shigenaga'14] / [高祖+'04, 犬伏+'09, 滝本+'15],
- **用例ごとに**人手による分析・計測が必要
 - 分析対象とした用例については高い信頼度
 - 新しい用例に対しては改めて分析・計測が必要



本研究: 大規模コーパスに基づく網羅的な分析

大規模コーパスに基づく分析

- 語順は直接観測可能な事象
 - コーパスから各語順の出現傾向を収集可能
 - 圧倒的に多くの用例に基づく網羅的検証
- 個別の事例の信頼性は低い
 - 個別事例から基本語順か判断するのは困難
 - 用例抽出の段階で誤りを含む可能性



夕飯を先輩に教わった店で食べた



夕飯を先輩に教わる



- 基本語順はコーパス中の出現割合と強く関係すると仮定
- 大規模コーパス(>100億文)中の確からしい用例のみ利用

二重目的語構文の基本語順の分析

- 分析に使用する用例の収集
- 動詞ごとの基本語順の分析
- 二格・ヲ格名詞を考慮した基本語順の分析
- まとめと今後の方向性

使用するコーパスと用例の収集方針

- 非常に規模の大きいコーパスから (>100億文)
 - Web上のテキスト集合を句点等をもとに文に分割
 - コピーページ対策のため重複文を削除し使用
- 構文的曖昧性の少ない用例だけを収集
 - 河原ら^[02]の手法で信頼性の高い部分を抽出 (精度98.3%)

先輩に教わった店で夕飯を食べた



先輩に教わる

店で夕飯を食べた

- ただし、項が複合名詞の場合は基本的に主辞のみ抽出
- ガ格や連体形等の扱いについては後述

分析対象とする動詞と収集結果

- 以下の条件を満たす動詞の用例を収集
 - a. 能動態で、形態素解析器JUMANの基本辞書に含まれる
 - b. ヲ格名詞、二格名詞がともにJUMAN辞書に含まれている用例の異なり数が500以上
 - c. ヲ格名詞、二格名詞を両方持つ用例の割合が5%以上
- 条件を満たした動詞は648個
 - 1動詞あたりの出現数の平均は約35万
 - ヲ格、二格をとともに持つ用例数の平均は約5.9万
 - ヲ格名詞、二格名詞がともにJUMAN辞書に含まれている用例の異なり数の平均は約1.0万

二重目的語構文の基本語順の分析

- 分析に使用する用例の収集
- **動詞ごとの基本語順の分析**
- 二格・ヲ格名詞を考慮した基本語順の分析
- まとめと今後の方向性

検証する仮説と分析方法

- a. 動詞によらず基本語順は「がにを」である [Hoji'85]
- b. 基本語順は動詞のタイプによって異なる [Matsuoka'03]
- c. 省略されにくい格は動詞の近くに出現しやすい

昨日に比べ肌寒さを感じます。

直前要素⇒意味を決める⇒省略されにくい

～に ... 感じる

<

～を ... 感じる

⇒ 基本語順:

～に ～を ... 感じる

～に ... 例える

>

～を ... 例える

⇒ 基本語順:

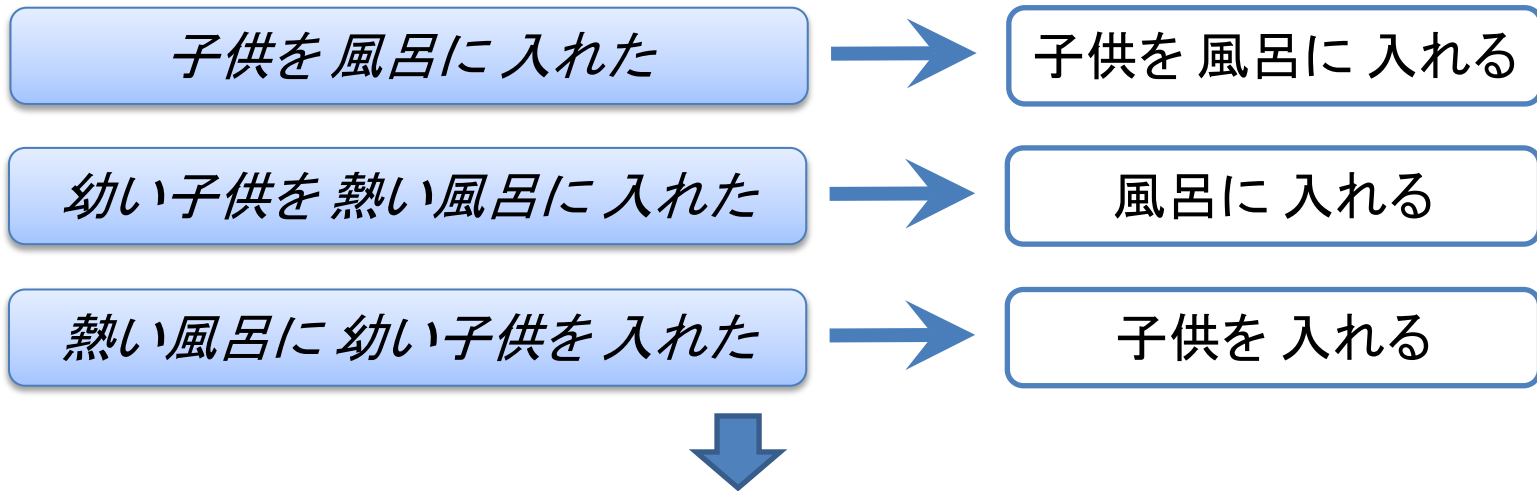
～を ～に ... 例える

野生動物に例えると...

- 前述の条件を満たす動詞ごとに以下を算出
 1. ヲ格と二格の一方のみ出現した用例に占める二格の割合(横軸)※
 2. ヲ格と二格が両方出現した用例に占める「をに」語順の割合(縦軸)

※ 河原ら[’02]の手法で収集した用例 そのままではバイアスが存在

- 基本的に曖昧性のない係り受けから収集



近い位置に出現する用例の方が集まりやすい！

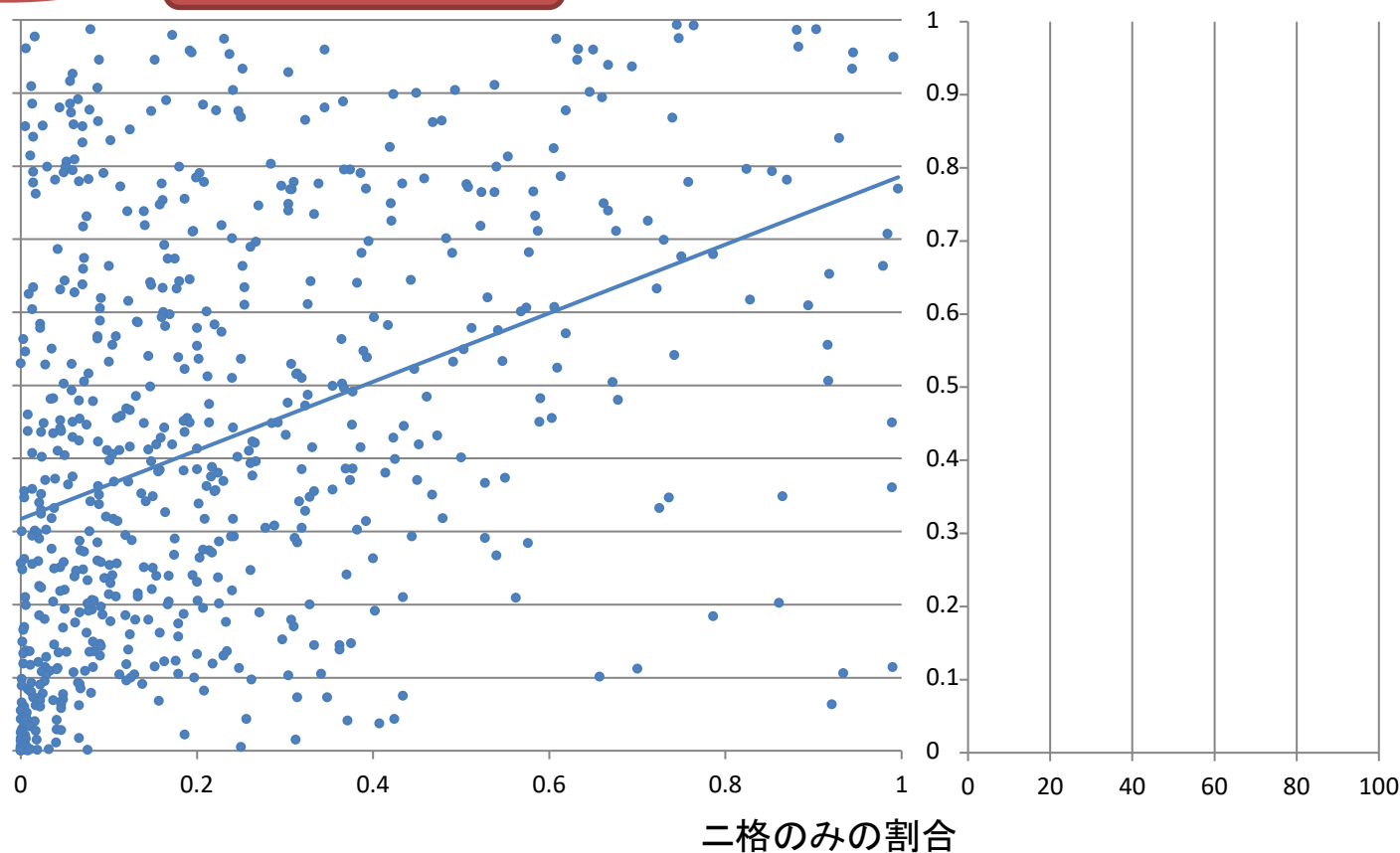
- [解決策] ヲ格とニ格の一方のみ出現した用例(=横軸)としてガ格も収集された用例のみを使用
- さらに、未格がある場合、連体形の動詞は収集対象から外す
(NAISTコーパスで検証⇒バイアスが解消されることを確認)

実験1: 格の出現率と語順

相関係数: 0.391

⇒ 弱い正の相関

「がをに」語順
の割合



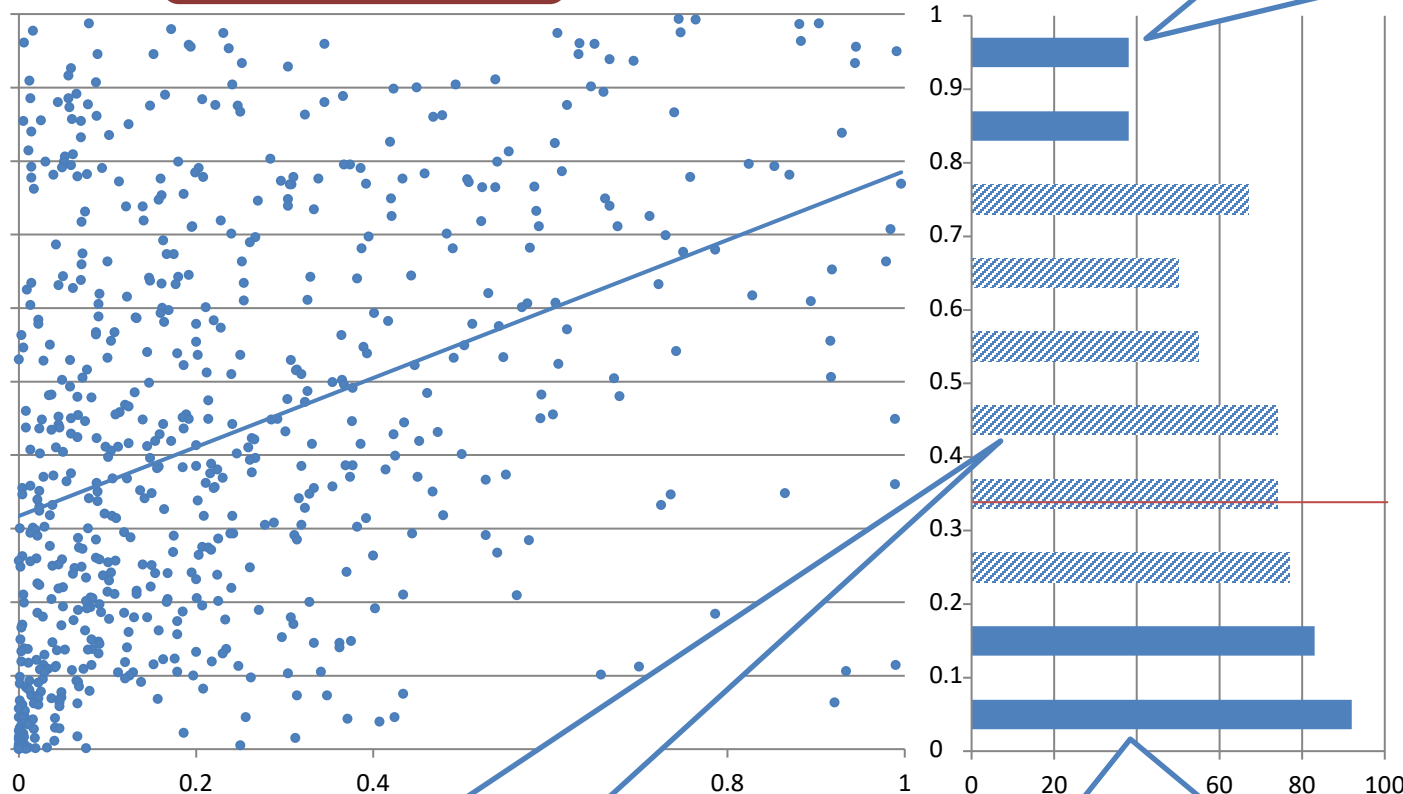
実験1: 格の出現率と語順

相関係数: 0.391

⇒ 弱い正の相関

「がをに」語順の割合

動詞の例と図中の縦軸、横軸の値
 例える:0.993:0.745, 見立てる:0.994:0.657
 代える:0.933:0.252, 生かす:0.916:0.056



動詞の例と図中の縦軸、横軸の値
 近付ける:0.487:0.326, 詰める:0.474:0.214
 上げる:0.448:0.026, 適用する:0.421:0.266

二格のみの割合

動詞の例と図中の縦軸、横軸の値
 発揮する:0.098:0.001, 感じる:0.043:0.256
 及ぼす:0.015:0.312, もたらす:0.000:0.000

平均: 0.328

実験1が示唆する内容

- a. 動詞によらず二重目的語構文の基本語順は「がにを」である [Hoji'85] ⇒ *No!*
 - ただし、全体で多い語順はガニヲ語順 (=67.2%) [Hoji'85]
- b. 基本語順は動詞のタイプによって異なる [Matsuoka'03] ⇒ *No?*
- c. 省略されにくい格は動詞の近くに出現しやすい ⇒ *Yes!*

基本語順と動詞タイプ

- 動詞タイプ: inchoative alternationを適用時に

- Showタイプ: 二格名詞が主語となる

彼に本を見せた ⇒ 彼が見た

- Passタイプ: ヲ格名詞が主語となる

本を彼に渡した ⇒ 本が渡った

- 基本語順との関係に関する仮説 [Matsuoka'03]

- Showタイプの動詞は「がにを」が基本語順

- Passタイプの動詞は「がをに」が基本語順

実験2: 動詞タイプごとの語順の割合

Show タイプ			Pass タイプ					
動詞	用例数	「がをに」率	動詞	用例数	「がをに」率	動詞	用例数	「がをに」率
知らせる	372927	0.522	戻す	146145	0.771	落とす	129406	0.351
預ける	77671	0.399	泊める	2551	0.748	漏らす	15193	0.332
言付ける	176	0.386	くるむ	1277	0.603	浮かべる	52038	0.255
論す	329	0.325	伝える	216113	0.522	向ける	319614	0.251
見せる	163172	0.301	乗せる	111757	0.496	残す	212591	0.238
被せる	23324	0.256	届ける	71428	0.491	埋める	50615	0.223
教える	144282	0.235	並べる	54949	0.481	混ぜる	39220	0.200
授ける	9294	0.186	返す	43045	0.448	当てる	203360	0.185
浴びせる	17714	0.177	ぶつける	66620	0.436	掛ける	164960	0.108
貸す	54359	0.118	付ける	529550	0.368	重ねる	74584	0.084
着せる	26791	0.113	渡す	154272	0.362	建てる	32203	0.069
平均 0.274						平均 0.365		

- 基本的に各タイプの動詞は[Koizumi+'04]が心理実験で使用した動詞を使用(out: はかせる, in: 知らせる&言付ける)
- 平均値の差を並べ替え検定で検定⇒有意差なし(p=0.177)
- 仮説bを支持しない(cf. [Miyamoto+'02, Koizumi+'04]も同様の結論)

二重目的語構文の基本語順の分析

- 分析に使用する用例の収集
- 動詞ごとの基本語順の分析
- **二格・ヲ格名詞を考慮した基本語順の分析**
- まとめと今後の方向性

検証する仮説

d. 基本語順は二格名詞の意味役割や有生性により異なる [Matsuoka'03, Ito'07]

- 二格名詞が有生性をもつ所有者(着点)を表す場合、有生性を持たない場所(着点)を表す場合よりも「がにを」語順をとりやすい

先生に本を返却した

⇔

本を学校に返却した

e. 省略されにくいヲ格名詞、二格名詞は動詞の近くに出現しやすい

脳に... 与える

<

影響を... 与える

⇒ 基本語順:

脳に影響を与える

身に... 付ける

>

能力を... 付ける

⇒ 基本語順:

能力を身につける

実験3: 二格のカテゴリと語順

- カテゴリはJUMAN 辞書のカテゴリ情報をもとに決定
 - 用例数が100万を超える8つのカテゴリが対象

カテゴリ	用例数	「がをに」率	出現頻度の高い名詞の例
場所-機能	1376990	0.499	下、横、外、中、方向、...
動物-部位	1483885	0.441	手、身、頭、肌、胸、...
人	5511281	0.387	友達、人、市長、私、先生、...
人工物-その他	2751008	0.372	パソコン、ファイル、風呂、写真、本、...
場所-施設	1618690	0.342	部屋、店、所、冷蔵庫、学校、...
場所-その他	2439188	0.341	場所、世界、サイト、位置、前面、...
数量	1100222	0.308	図、何、表、半分、値、...
抽象物	10219318	0.307	ブログ、心、リスト、視野、元、...
合計	26500582	0.353	

- カテゴリにより「がをに」語順の割合が異なることを確認
 - ただし、二格が有生名詞(cf. 人)の場合「がにを」語順をとりやすいという仮説と合致しない
 - 意味役割を限定できていないことが原因である可能性

実験4: 二格の意味役割と語順

- [滝本+'15]を参考に下記を抽出 (e.g.「本を学校に返却」「先生に本を返却」)
 - 二格名詞のJUMAN辞書におけるカテゴリが『人工物-その他』
 - 二格名詞のカテゴリが『人』である用例数と、『場所-施設』である用例数がいずれも100以上である126動詞
- 二格名詞のカテゴリごとに語順の割合を調査
 - 語順に有意な差があった動詞94個中、『人』である場合の方が「がにを」語順をとりやすい動詞は64個、その逆は動詞は30個
 - 仮説dを支持する結果

動詞	「がにを」率 二格が『人』 『場所-施設』	関連する出現頻度の高い用例の抜粋とその頻度
据える	0.852 > 0.203	アクセルを <u>主役に</u> -:8、ターゲットを <u>女性に</u> -:6、ホテルにカメラを -:4
展示する	0.851 > 0.359	館に美術品を -:46、館に資料を -:46、製品を <u>一同に</u> -:39、商品を <u>一同に</u> -:26
返還する	0.915 > 0.491	許可証を市長に -:253、登録証を市長に -:149、営業店にレンタカーを -:51
運ぶ	0.244 < 0.577	荷物を部屋に -:250、荷物を部屋に -:188、子供に餌を -:92、客に料理を -:73
落とす	0.173 < 0.595	携帯をトイレに -:486、電話をトイレに -:282、トイレに携帯を -:168、私に爆弾を -:16
郵送する	0.224 < 0.664	合格者に書類を -:180、書類を自宅に -:156、用紙を客に -:61、合格者に証書を -:48

検証する仮説

d. 基本語順は二格名詞の意味役割や有生性により異なる [Matsuoka'03, Ito'07]

- 二格名詞が有生性をもつ所有者(着点)を表す場合、有生性を持たない場所(着点)を表す場合よりも「がにを」語順をとりやすい

先生に本を返却した

⇔

本を学校に返却した

e. 省略されにくいヲ格名詞、二格名詞は動詞の近くに出現しやすい

脳に... 与える

<

影響を... 与える

⇒ 基本語順:

脳に影響を与える

身に... 付ける

>

能力を... 付ける

⇒ 基本語順:

能力を身につける

ヲ格名詞、二格名詞、動詞の組み合わせごとの「がをに」語順の割合の調査

- ヲ格名詞と動詞、二格名詞と動詞、それぞれの共起度合いと、「がをに」語順の関係を調査

- 名詞と動詞の共起度合いの尺度には、以下で定義される正規化自己相互情報量(NPMI)を使用(常に共起⇒1、独立⇒0、共起しない⇒-1)

$$\text{NPMI}(n_{\pm}, v) = \frac{\text{PMI}(n_{\pm}, v)}{-\log(p(n_{\pm}, v))} \quad \text{ただし、PMI}(n_{\pm}, v) = \log \frac{p(n_{\pm}, v)}{p(n_{\pm})p(v)}$$

- 二格名詞と動詞のNPMIとヲ格名詞と動詞のNPMIの差を算出
 - 仮説eが正しいならば二格名詞と動詞のNPMIの方が大きな値となる場合、「がをに」語順の割合が大きくなるはず
- 500回以上出現した組み合わせ2417個を調査

慣用表現等への対処

- 基本的に慣用表現はまとまって出現
 - 全用例を人手でチェック
 - Yes: 「～を手に入れる」、「世界を股にかける」
 - ???: 「声をかける」、「視線を向ける」
- 機能動詞構文も影響する可能性もあるが保留
 - Yes: 「実行に移す」、???: 「影響を与える」
- その他、不適切な用例は人手で削除 (2417⇒2302)
 - 「～を友達に知らせる」: SNSの定型表現
 - 「～を最大限に活用する」

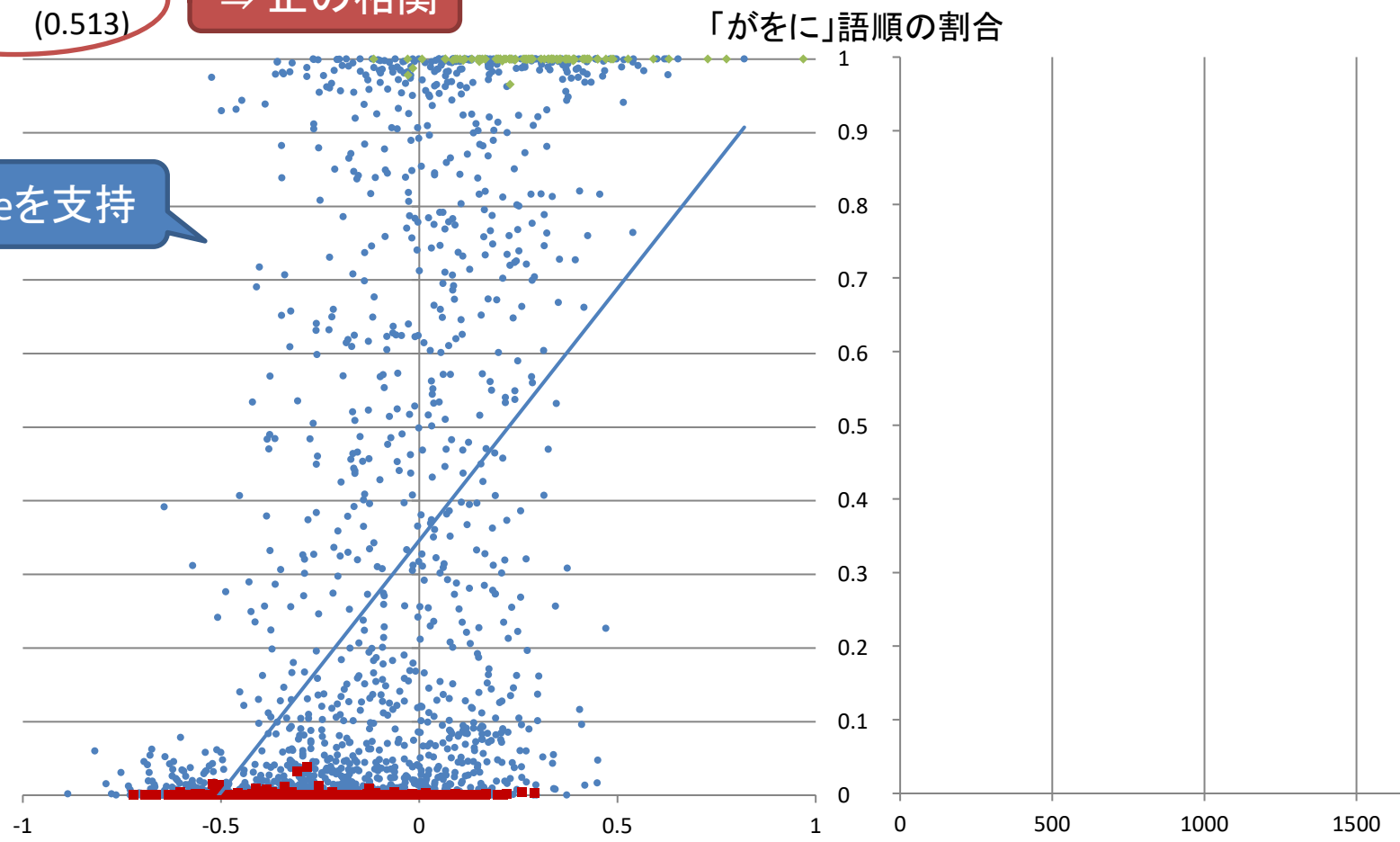
実験5: ヲ格名詞、二格名詞、動詞の組み合わせごとのNPMIの差と語順の関係

1814例、ヲ二: 404例、ニヲ: 84例

相関係数: 0.567
(0.513)

⇒ 正の相関

仮説eを支持



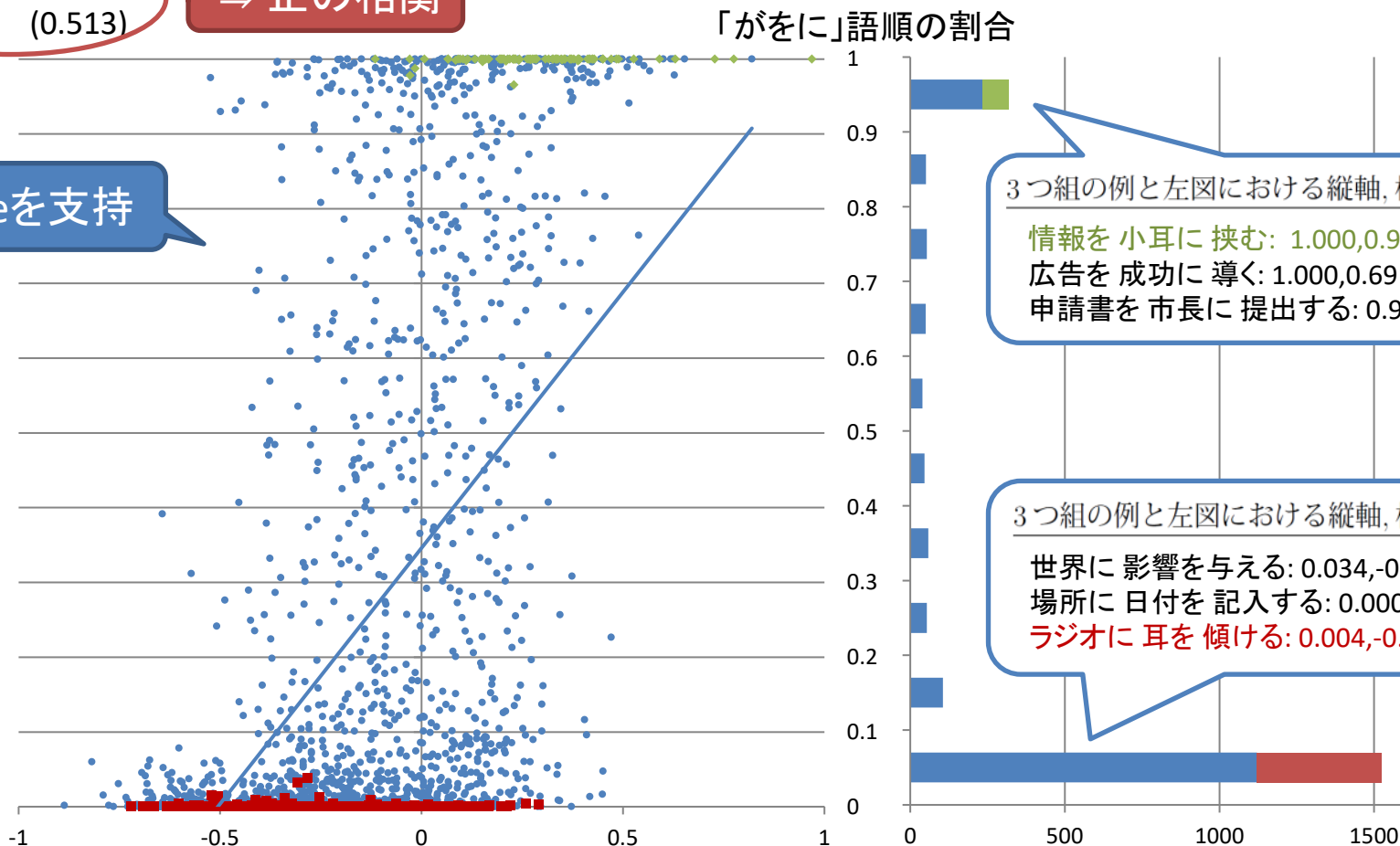
実験5: ヲ格名詞、二格名詞、動詞の組み合わせごとのNPMIの差と語順の関係

1814例、ヲ二: 404例、ニヲ: 84例

相関係数: 0.567
(0.513)

⇒ 正の相関

仮説eを支持



二重目的語構文の基本語順の分析

- 分析に使用する用例の収集
- 動詞ごとの基本語順の分析
- 二格・ヲ格名詞を考慮した基本語順の分析
- **まとめと今後の方向性**

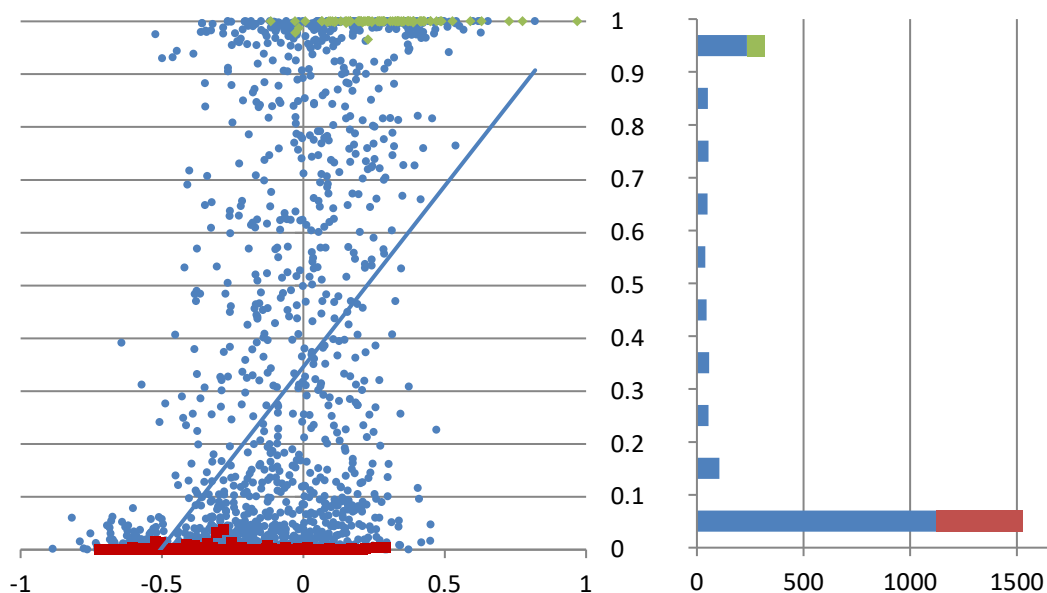
まとめ: 大規模コーパスに基づく日本語 二重目的語構文の基本語順に関する考察

- a. 動詞によらず二重目的語構文の基本語順は「がにを」
[Hoji'85] ⇒ **No!** (ただし、全体で多いのは「がにを」語順=67.2%)
- b. 基本語順は動詞タイプにより異なる [Matsuoka'03] ⇒ **No!**
- c. 省略されにくい格は動詞の近くに出現しやすい ⇒ **Yes!**
- d. 基本語順は二格名詞の意味役割や有生性により異なる
[Matsuoka'03, Ito'07] ⇒ **Yes!**
- e. 省略されにくいヲ格名詞、二格名詞は動詞の近くに出現
しやすい ⇒ **Yes!**

今後の方向性: 語順と意味

- 実験5の結果から考えられる仮説
 1. 基本語順は名詞・動詞の組み合わせで決まる
 2. 基本語順は名詞・動詞の意味関係で決まる [中本+'06]

- 2が正しいなら
 - 類義語に置き換えても基本語順は不変のはず
 - ⇒ データで検証
 - 意味解析の手掛かりとして活用



今日の話

- 言語現象に対する各種アプローチ
 - 言語学からのアプローチ
 - 脳科学・認知科学からのアプローチ
 - 自然言語処理(NLP)/計算言語学(CL)からのアプローチ
- 日本語二重目的語構文の基本語順の分析
 - 日本語二重目的語構文と既存のアプローチ
 - 大規模コーパスに基づく基本語順の分析

まとめ

- 自然言語処理(NLP/CL)と言語学
 - 言語現象の分析でNLP/CLの技術が有用な場合も多い
 - 特に多数の組み合わせを考慮すべき現象では、まずコーパスに基づく分析で仮説を検証するのも有力な方法

	コスト	大規模化	客観性	信頼性	直接性
NLP/CL	😊	😊	😊	😞	😞
言語学	😞	😞	😞	😊	😞
脳科学	😞	😞	😊	😞	😊

- 大規模コーパスに基づく二重目的語構文の分析
 - 「基本語順は二格名詞の意味役割や有生性により異なる」
 - 「省略されにくい格は動詞の近くに出現しやすい」等を確認