

# Language Models as an Alternative Evaluator of Word Order Hypotheses: A Case Study in Japanese

---

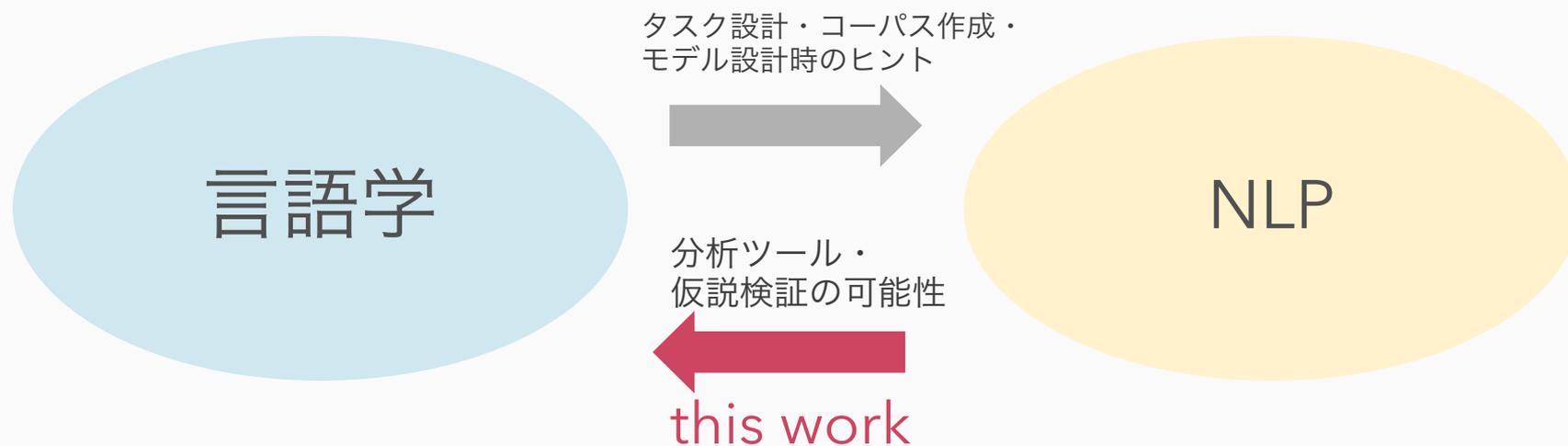
Tatsuki Kuribayashi, Takumi Ito, Jun Suzuki, Kentaro Inui

読む人: 東北大学D1 栗林樹生

## 経緯

もともと神藤さんが本論文を含めて二本紹介してくださるということでしたが、著者が運営にいたるため、著者が発表する運びとなりました

# NLPのツールが言語研究に役立つ...?



言語モデルを語順分析ツールとして売り込む

# 知りたい対象: 語順

---

- 言語の線状性: 言語における重力

同時に一文字(音)以上伝達できません

※手話除く

→ time

# 知りたい対象: 語順

- 言語の線状性: 言語における重力

同時に一文字(音)以上伝達できません

→ time

※手話除く

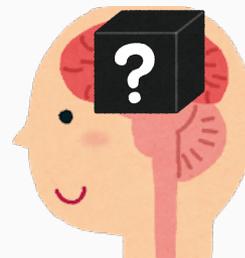
- どのように並べるか？

- 基本語順 (母語話者の処理負荷が低い) の存在を仮定

- 文法で説明できる部分 (SVO, SOV...) ・ **いわゆる直観**

「影響を質に与える」よりも「質に影響を与える」

「鳥に人を例える」よりも「人を鳥に例える」



# 知りたい対象: 語順

- 言語の線状性: 言語における重力

同時に一文字 (音) 以上伝達できません

→ time

※手話除く

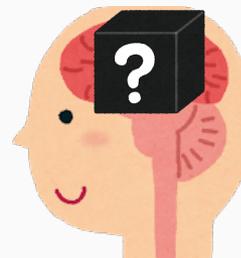
- どのように並べるか？

- 基本語順 (母語話者の処理負荷が低い) の存在を仮定

- 文法で説明できる部分 (SVO, SOV...) ・ **いわゆる直観**

「影響を質に与える」よりも「質に影響を与える」

「鳥に人を例える」よりも「人を鳥に例える」



- 基本語順規則に対して様々な仮説

- 例. 二重目的語の基本語順は動詞のタイプによって異なる

分かんると...



日本語への理解  
言語類型研究  
思考の順序

言語学・認知科学

言語教育  
ライティング  
支援ツール

教育・応用

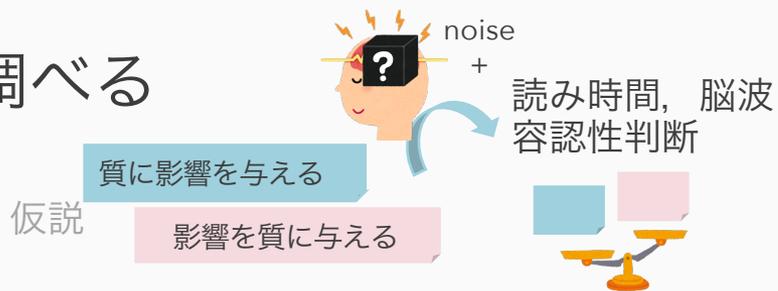
FAQ. 文脈に応じて好まれる  
語順が変わるよね？

-- この研究では、文脈非依存に支配的な語順 (基本語順) の傾向に焦点を絞る

# どうやって仮説を検証する？

- 1. ヒトを調べる

[心理言語学]



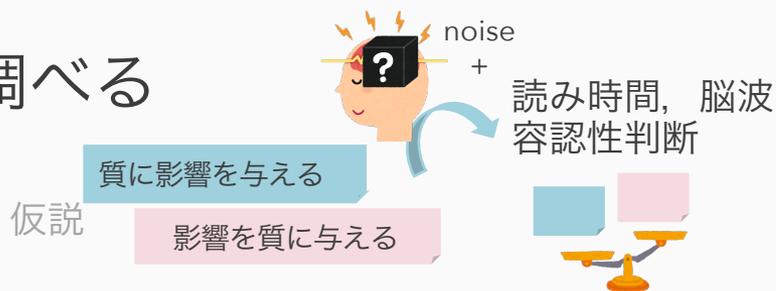
直接的だが高コスト

偏りのないヒトを集める, 実験設定の統制  
アイトラッカーなど異様に高額

# どうやって仮説を検証する？

## ● 1. ヒトを調べる

[心理言語学]



直接的だが高コスト

偏りのないヒトを集める, 実験設定の統制  
アイトラッカーなど異様に高額

## ● 2. コーパス頻度を調べる

[コーパス言語学全般][Sasano+, 2016]

仮定: 基本語順は産出されやすい

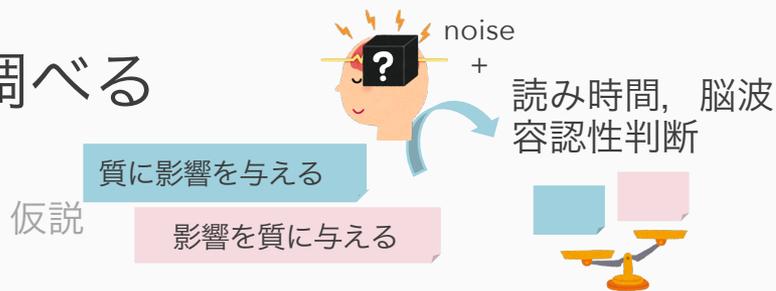


間接的だが低コスト

# どうやって仮説を検証する？

- 1. ヒトを調べる

[心理言語学]



直接的だが高コスト

偏りのないヒトを集める, 実験設定の統制  
アイトラッカーなど異様に高額

- 2. コーパス頻度を調べる

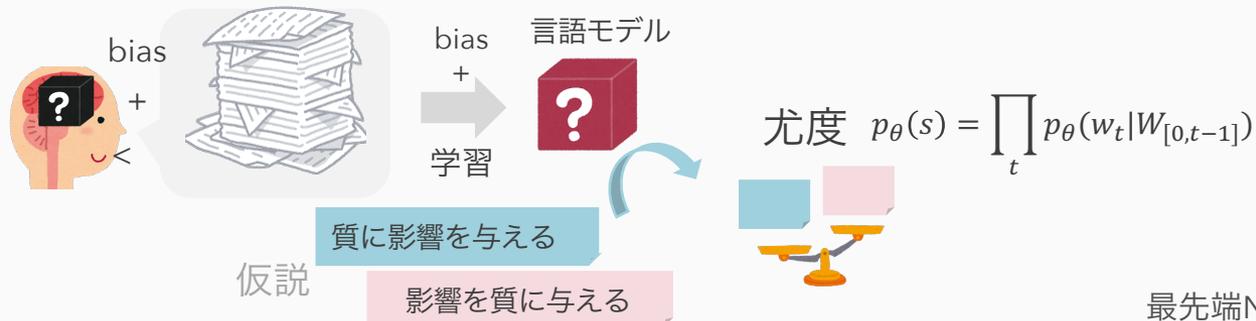
[コーパス言語学全般][Sasano+, 2016]

仮定: 基本語順は産出されやすい



間接的だが低コスト

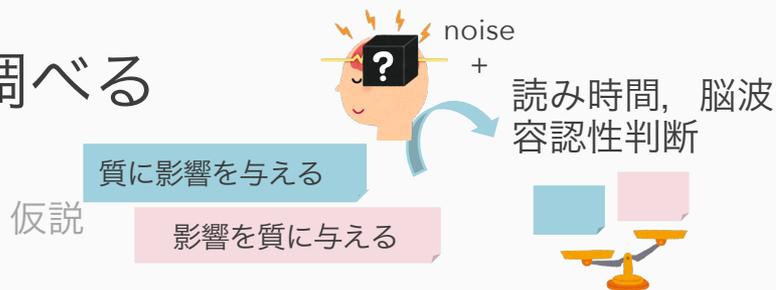
- 3. 言語モデル尤度で比較する (本研究)



# どうやって仮説を検証する？

## ● 1. ヒトを調べる

[心理言語学]



直接的だが高コスト

偏りのないヒトを集める, 実験設定の統制  
アイトラッカーなど異様に高額

## ● 2. コーパス頻度を調べる

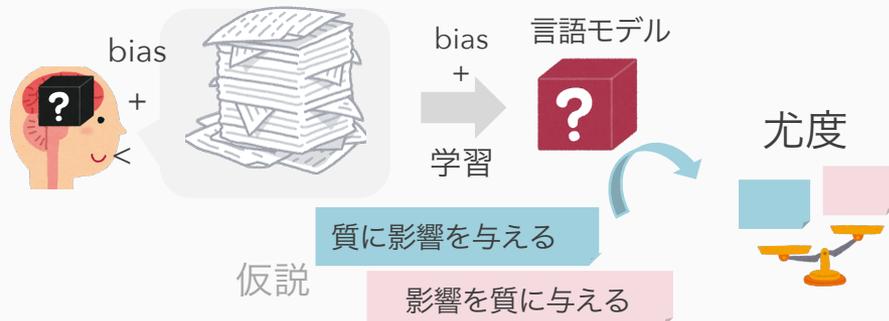
[コーパス言語学全般][Sasano+, 2016]

仮定: 基本語順は産出されやすい



間接的だが低コスト

## ● 3. 言語モデル尤度で比較する (本研究)



上2つの手段よりも

- 実現しやすく

- 経験的に同じ結果が導かれる

# 売り込みポイント1: 実験コスト

## ■ 時に解析器の用意が大変

ドメイン間の傾向を分析したい (話し言葉, 学習者の文, 詞...)  
 マイナーな言語  
 同定の難しい現象 (格助詞の省略と語順の関係を調べたい)

脳波

斤

■ 言語モデル学習には  
生コーパスがあれば良い

■ 同定は難しいが作例は  
簡単なケースがある

格助詞が省略された文を  
作成するのは簡単

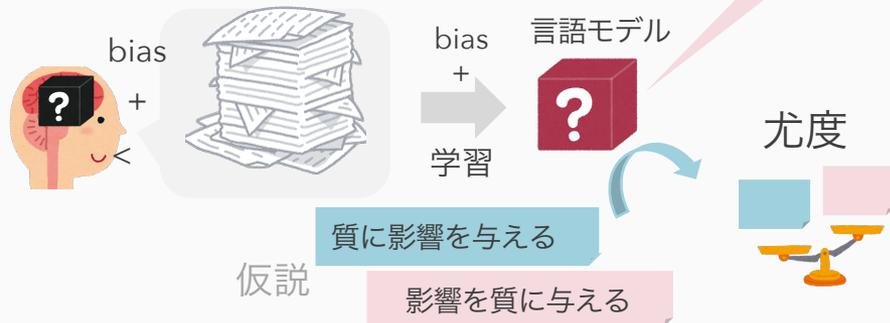
## ● 2. コーパス頻度を調

[コーパス言語学全般][Sasano+, 2016]

同定: 基本語順は産出



## ● 3. 言語モデル尤度で比較する (本研究)



上2つの手段よりも

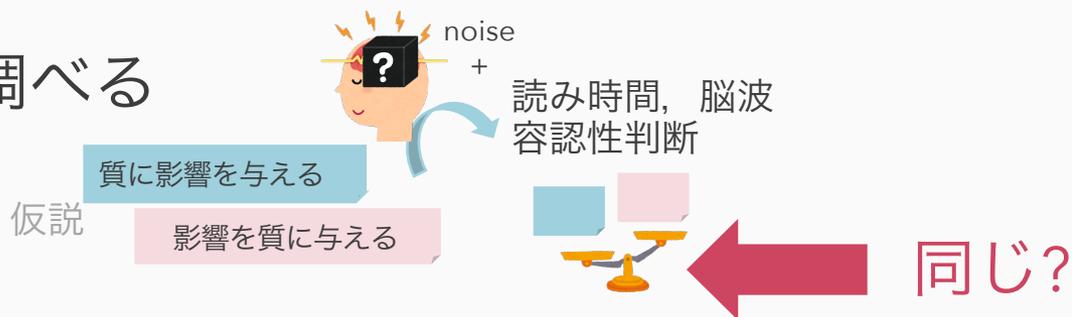
- 実現しやすく

- 経験的に同じ結果が導かれる

# 売り込みポイント2: 経験的に同じ結論が導かれる

## ● 1. ヒトを調べる

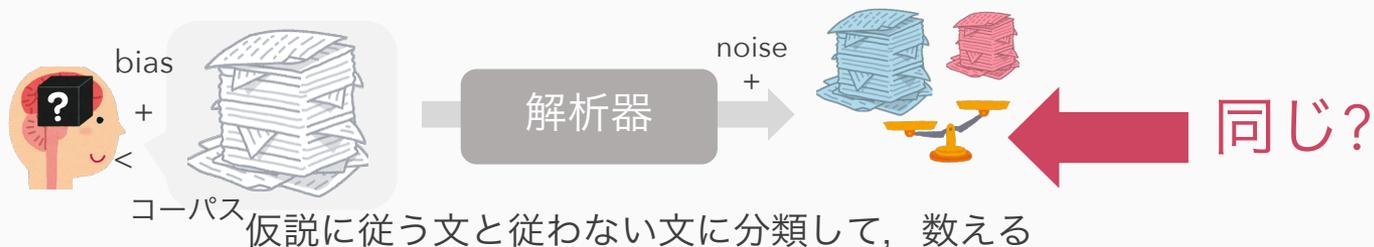
[心理言語学]



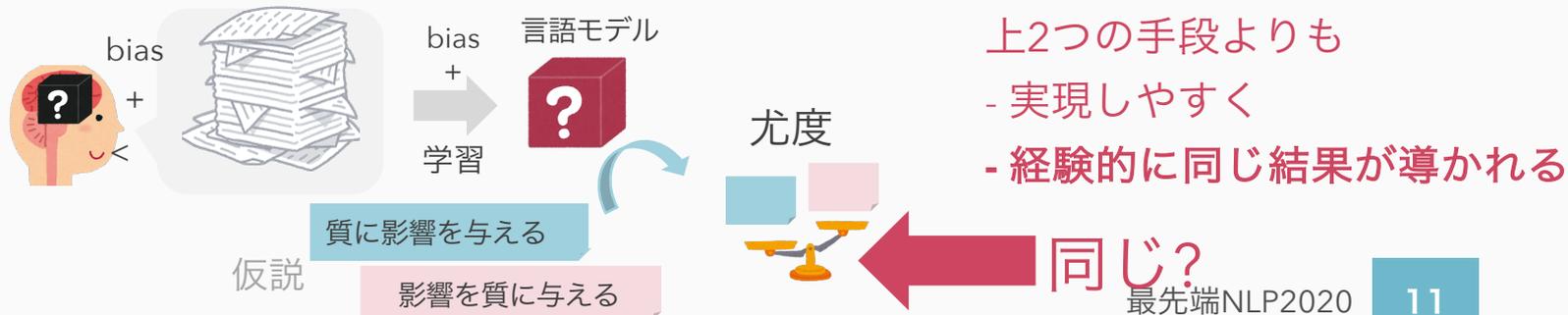
## ● 2. コーパス頻度を調べる

[コーパス言語学全般][Sasano+, 2016]

仮定: 基本語順は産出されやすい

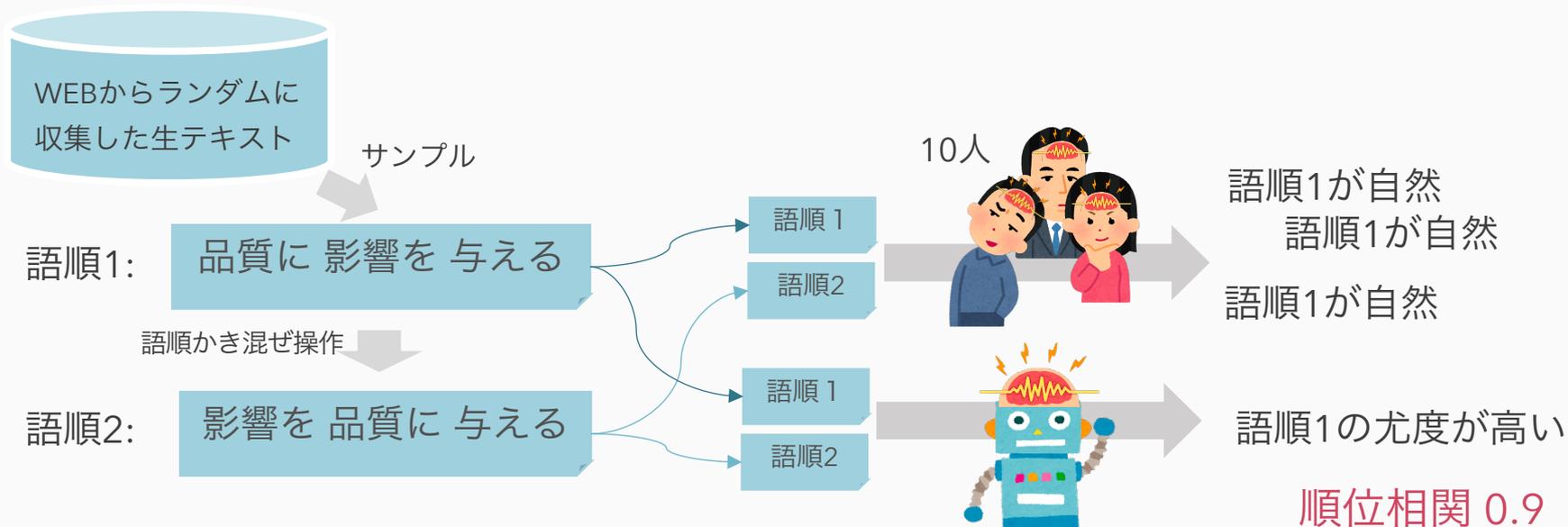


## ● 3. 言語モデル尤度で比較する (本研究)



# 実験1: 人間の選好との相関

- Yahoo!クラウドソーシングを利用 (計756人が参加)



以下の条件を満たす2.6k 文ペアを使用

- どちらの語順も非文でない(ワーカーによる判断)
- ワーカー10人中9人以上の選好が一致

ヒトと言語モデルの  
選好が非常に近い

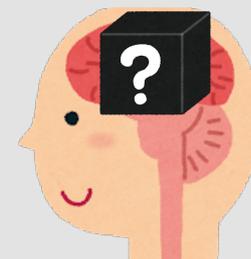
# 補足: ヒトの認知負荷のモデル化

- 近年, 言語モデルサプライザルとヒトの認知負荷の間に強い関係があることが共通認識に [Smith and Levy, 2013] [Merks and Frank, 2020] [Goodkind and Bicknell, 2018] [Wilcox+, 2020]
  - 読み時間・脳活動をどれぐらいモデル化できるかで議論

この単語の  
認知負荷を計算したい

(続きの) エントロピー

$$H(t) = - \sum_{v \in V} p(v|W_{[0,t]}) \log_2 p(v|W_{[0,t]})$$



...  $W_{t-2}$      $W_{t-1}$      $W_t$      $W_{t+1}$      $W_{t+2}$     ...

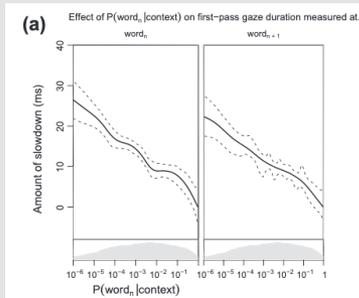
サプライザル (predictability)

$$-\log_2 p(w_t|W_{[0,t-1]})$$

エントロピーの変化

$$H(t) - H(t-1)$$

lead to more processing effort. For instance, in psycholinguistics it is common to take reading times as a measure of word processing difficulty and the positive correlation between reading time and surprisal has firmly been established (Hale, 2001; Levy, 2008; Mitchell and Keller, 2010; Monsalve et al., 2012; Smith and Levy, 2013; Hahn and Keller, 2016) with Goodkind and Bicknell (2018) recently showing



[Smith and Levy, 2013]

[Wilcox+, 2020]

(英語における実験によって...英語母語話者にしか興味ないの...)

# 補足: ヒトの認知負荷のモデル化

- 「(単方向) 言語モデルで計算した文尤度が高い」  
言い換えれば「文全体の累積サプライズルが小さい」

$$p(s) = \prod_t p(w_t | W_{[0,t-1]}) \text{ が大きい}$$

$$\log_2 \frac{1}{p(s)} = \sum_t \underbrace{-\log_2 p(w_t | W_{[0,t-1]})}_{\text{各単語のサプライズル}} \text{ が小さい}$$

あくまでクラウドワーカーの内省

- ヒトの語順容認度と言語モデルサプライズルに関係があることを語順が自由な言語において初めて示唆
  - 言語モデル尤度 (累積サプライズル) とヒトの語順選好に強い相関 (実験1より)

# 実験2: コーパス頻度に基づく検証との一貫

対象	仮説	既存の検証結果	言語モデル
二重目的語	動詞によらず基本語順は「にを」である	棄却	棄却
	基本語順は動詞のタイプによって異なる	棄却	棄却
	省略されにくい格は基本語順において動詞の近くに位置する	支持	支持
	基本語順は二格名詞の意味役割や有生性によって異なる	支持	支持
	対象の動詞と高頻度に共起するヲ格, 二格名詞は動詞の近くに位置する	支持	支持
副詞	副詞のタイプにより基本生起位置は異なる	支持	支持
主語	主語は時間・場所を表す格よりも後ろにくる	支持	支持
場所	場所を表す格は時を表す格より後ろ, 主語よりも前に位置する	支持	支持
時	時間を表す格は場所, 主語よりも前に位置する	支持	支持
(一般)	長い句が短い句よりも前に来る	支持	支持

各仮説の詳細は省略

一致

# 補足: 語順の計算的な説明

対象	仮説
二重目的語	動詞によらず基本語順は「にを」である -> なさそう
	基本語順は動詞のタイプによって異なる -> 異ならなさそう
	省略されにくい格は基本語順において動詞の近くに位置する
	基本語順は二格名詞の意味役割や有生性によって異なる
	対象の動詞と高頻度に共起するヲ格, 二格名詞は動詞の近くに位置する
副詞	副詞のタイプにより基本生起位置は異なる
主語	主語は時間・場所を表す格よりも後ろにくる
場所	場所を表す格は時を表す格より後ろ, 主語よりも前に位置する
時	時間を表す格は場所, 主語よりも前に位置する
(一般)	長い句が短い句よりも前に来る

コーパス上で学習した言語モデルと情報量の観点で統一的に記述できる

$$\operatorname{argmax}_{\theta} \sum_t \log_2 p_{\theta}(w_t | W_{[0,t-1]})$$



驚かないような語順〜



$$\operatorname{argmin}_{w \in \text{可能な語順}} \sum_t -\log_2 p_{\theta}(w_t | W_{[0,t-1]})$$

# 分析: 内省による仮説に経験的な示唆を

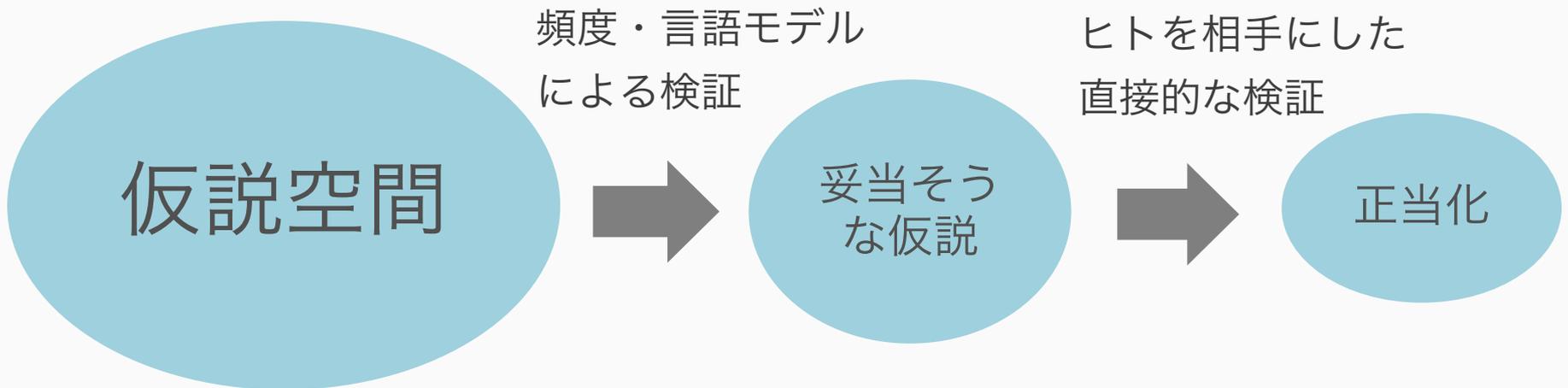
対象	仮説	既存の検証結果	言語モデル
二重目的語	動詞によらず基本語順は「にを」である	棄却	棄却
	基本語順は動詞のタイプによって異なる	棄却	棄却
	省略されにくい格は基本語順において動詞の近くに位置する	支持	支持
	基本語順は二格名詞の意味役割や有生性によって異なる	支持	支持
	対象の動詞と高頻度に共起するヲ格, 二格名詞は動詞の近くに位置する	支持	支持
副詞	副詞のタイプにより基本生起位置は異なる	支持	支持
主語	主語は時間・場所を表す格よりも後ろにくる	支持	支持
場所	場所を表す格は時を表す格より後ろ, 主語よりも前に位置する	支持	支持
時	時間を表す格は場所, 主語よりも前に位置する	支持	支持
(一般)	長い句が短い句よりも前に来る	支持	支持
とりたて	基本語順で前に来る格ほど主題化されて文頭に移動しやすい	-	支持
	目的語の主題化されやすさは動詞の「をに」率に依存	-	支持
	とりたてられた格の移動しやすさは副助詞と格に依存	-	支持

各仮説の  
詳細は省略

分析

# 結論

- 言語モデル尤度で日本語基本語順の傾向を分析する方法論について妥当性を支持する結果を提示した
  - 検証したい仮説によっては、より現実的な方法



# 補足: 産出されやすい言語 (コーパス頻度) と基本語順が異なる言語もあるらしい [Koizumi+ 2014]

- カクチケル語では、文法的基本語順がVOS語順であり、彼らにとっては文処理の負荷が低い
- しかし産出頻度はSVO語順が高い
- 出来事を認識する際の順序はSOの順



# 関連研究

- Futrell, R., & Levy, R. (2019). Do RNNs learn human-like abstract word order preferences?. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2019* (pp. 50-59).
- Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302-319.
- Goodkind, A., & Bicknell, K. (2018, January). Predictive power of word surprisal for reading times is a linear function of language model quality. In *Proceedings of the 8th workshop on cognitive modeling and computational linguistics (CMCL 2018)* (pp. 10-18).
- Wilcox, E. G., Gauthier, J., Hu, J., Qian, P., & Levy, R. (2020). On the Predictive Power of Neural Language Models for Human Real-Time Comprehension Behavior. *arXiv preprint arXiv:2006.01912*.
- Merx, D., & Frank, S. L. (2020). Comparing Transformers and RNNs on predicting human sentence processing data. *arXiv preprint arXiv:2005.09471*.
- Sasano, R., & Okumura, M. (2016, August). A Corpus-Based Analysis of Canonical Word Order of Japanese Double Object Constructions. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 2236-2244).
- 小泉政利 「言語の語順と思考の順序：カクチケル・マヤ語からみた人間の文処理メカニズム」, 日本語教育学講座講演会, 2014年12月4日, 名古屋大学. (招待講演)