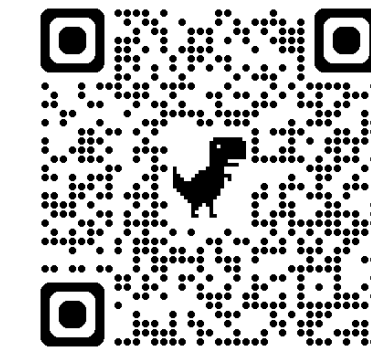




言語の起源はコミュニケーションか？：計算心理言語学の観点から

梶川康平

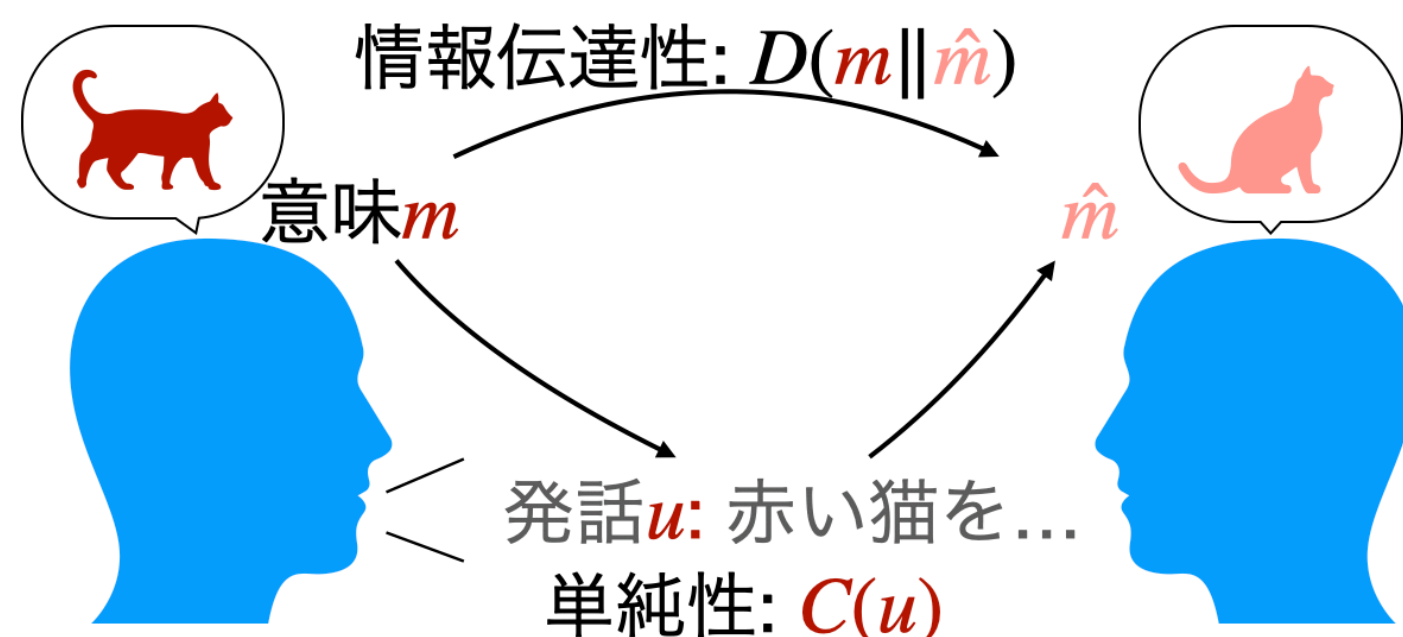
東京大学大学院 言語情報科学専攻修士2年・国立国語研究所



言語はコミュニケーションにおいて効率的か？

自然言語にはなぜは普遍性が存在するのか？

- 効率的なコミュニケーションを実現した結果だから [Jaeger & Tily, 2011; Kemp et al., 2018; Gibson et al., 2019]
- 認知制約のもと、伝わる意図・情報が最大化されている (informative) 一方、産出・理解のコストが最小化されている (simple) 状況
 - 情報伝達性と単純性のトレードオフのもとでの最適化



なぜいまこの研究？

- 情報理論の進展@心理言語学
 - コミュニケーションの数理理論と人間の言語使用の関係が徐々に明らかに
 - 言語使用それ自体が効率的、合理的
- 汎用的な言語処理モデルの登場
- 計算機上でのシミュレーションという方法論の確立
 - 「反事実的 (counterfactual)」な状況の検討が可能

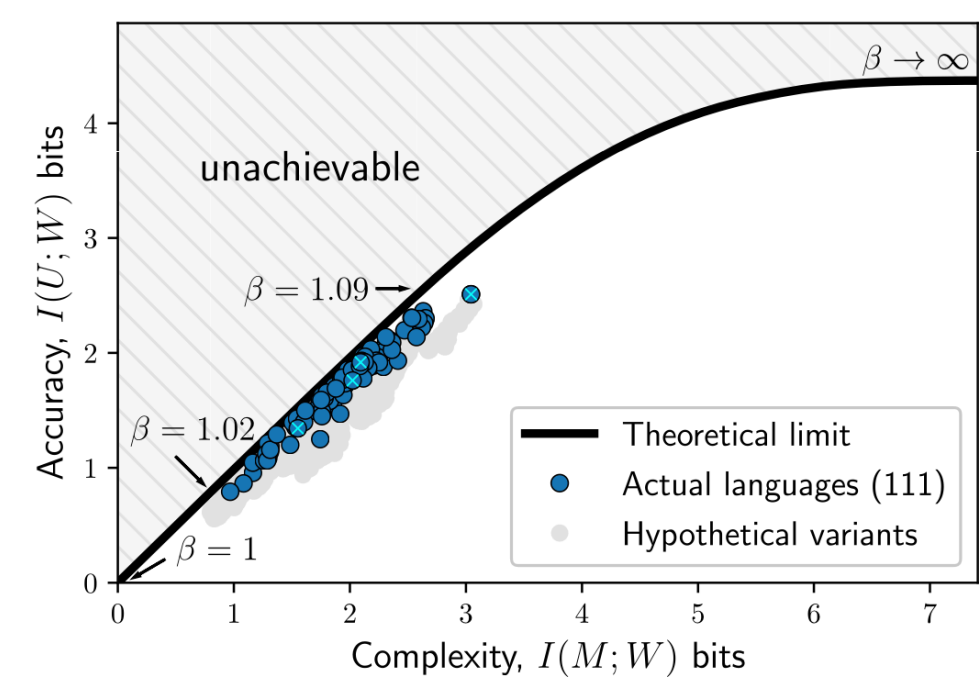
これまで何がわかっている？

語彙

- Wordform [Zipf, 1949; Piantadosi et al., 2011; Mahowald et al., 2018]
- Semantic categorization (色 [Regier et al., 2007; Zaslavsky et al., 2018]、親族名称 [Kemp & Regier, 2012]、数 [Xu et al., 2020; Denic & Szymanik, 2024]、文法標識 [Mollica et al., 2021]、量子化 [Steinert-Threlkeld, 2021]、などなど)

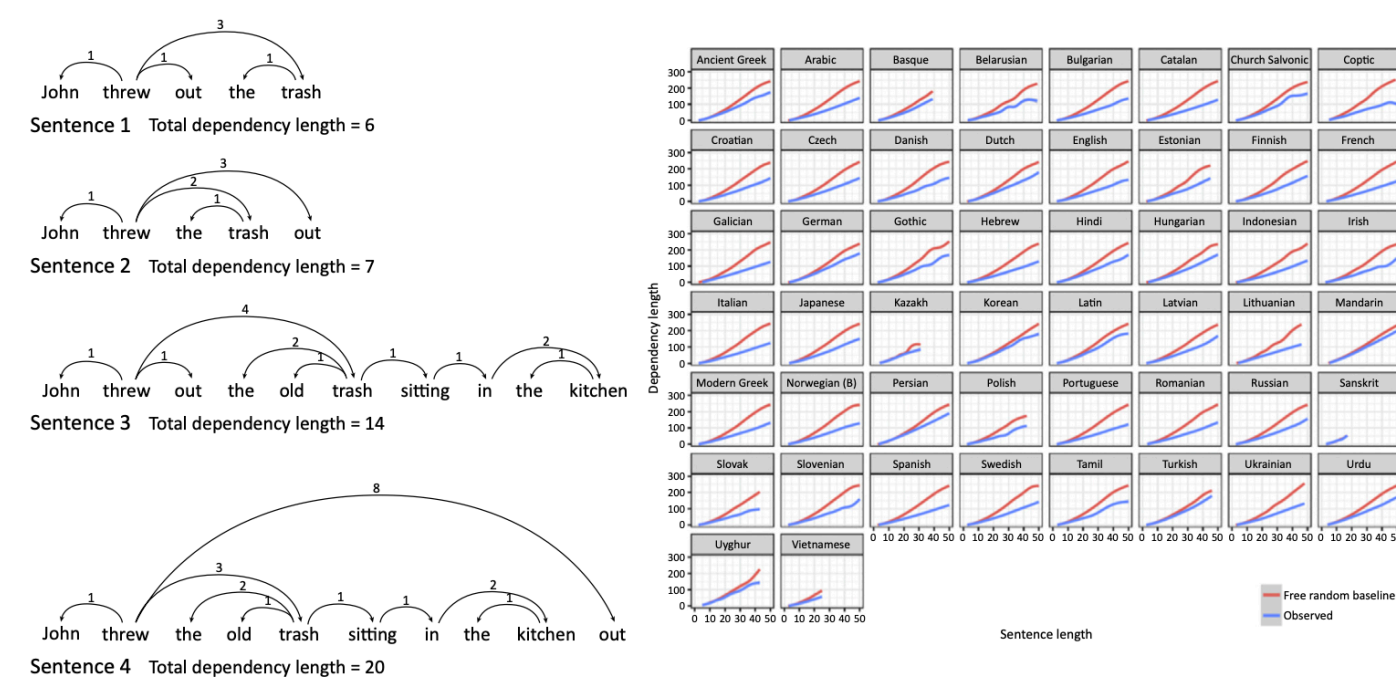
統語

- 構成性 [Kirby et al., 2015; Futrell & Hahn, 2024]
- 語順 [Gildea & Jaeger, 2015; Futrell et al., 2020; Hahn et al., 2021; Clark et al., 2023]
- 構造依存性 [Kajikawa et al., 2024] / 自由語順と項省略の文法相関 [梶川ら, 2025]



Zaslavsky et al. (2018), Fig. 3 より

色の名前は、コミュニケーションの最適解



Gibson et al. (2019), Fig. 4 より

語順は、依存関係の距離が短くなるようになっている

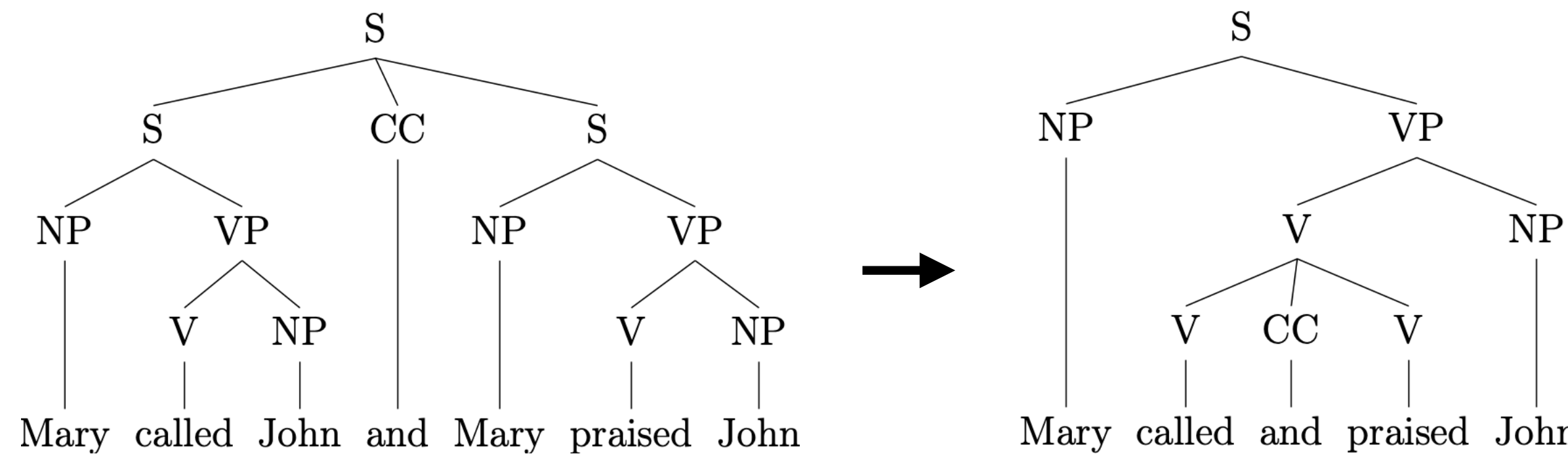
何がおもしろい？

- 普遍性（多様性）を領域一般の認知能力に帰する
- 言語が使用の影響を受けてるってアタリマエでは？
 - その当たり前のことすら、場合によっては（根拠なく）否定されている
 - 何が領域一般の認知能力で説明できるのかはっきりさせたい
- 「最適解」とかいけど、人間ってそんなに合理的？
 - 合理的ではない部分があるとしても、それが「どこでどの程度か？」ということを知るためには、合理的な理論が役にたつ (see also Griffiths (2024))

Case Study1: 構造依存性 [Kajikawa et al., 2024]

構造依存性は本当に領域固有のものか？

- コミュニケーションは付随的なもの (epiphenomena)？言語の本質ではない？ [Chomsky, 2002; Hauser et al., 2002]
- 構造依存性 (structure dependence) は本質かつ領域固有のもの？ [Hauser et al., 2002; Chomsky, 2005; Everaert et al., 2015, Berwick & Chomsky, 2016]
 - たとえば、等位接続 [Chomsky, 1957, 1955; Ross, 1967]



単純性と情報伝達性のトレードオフ

単純性・情報伝達性と、コミュニケーションの効率性の定量化：

$$\text{predictability (simplicity)} := -H(\mathcal{U}) = \sum_{u \in \mathcal{U}} p(u) \log p(u)$$

$$\text{parsability (informativeness)} := -H(\mathcal{T} | \mathcal{U}) = \sum_{t \in \mathcal{T}} \sum_{u \in \mathcal{U}} p(t, u) \log p(t | u)$$

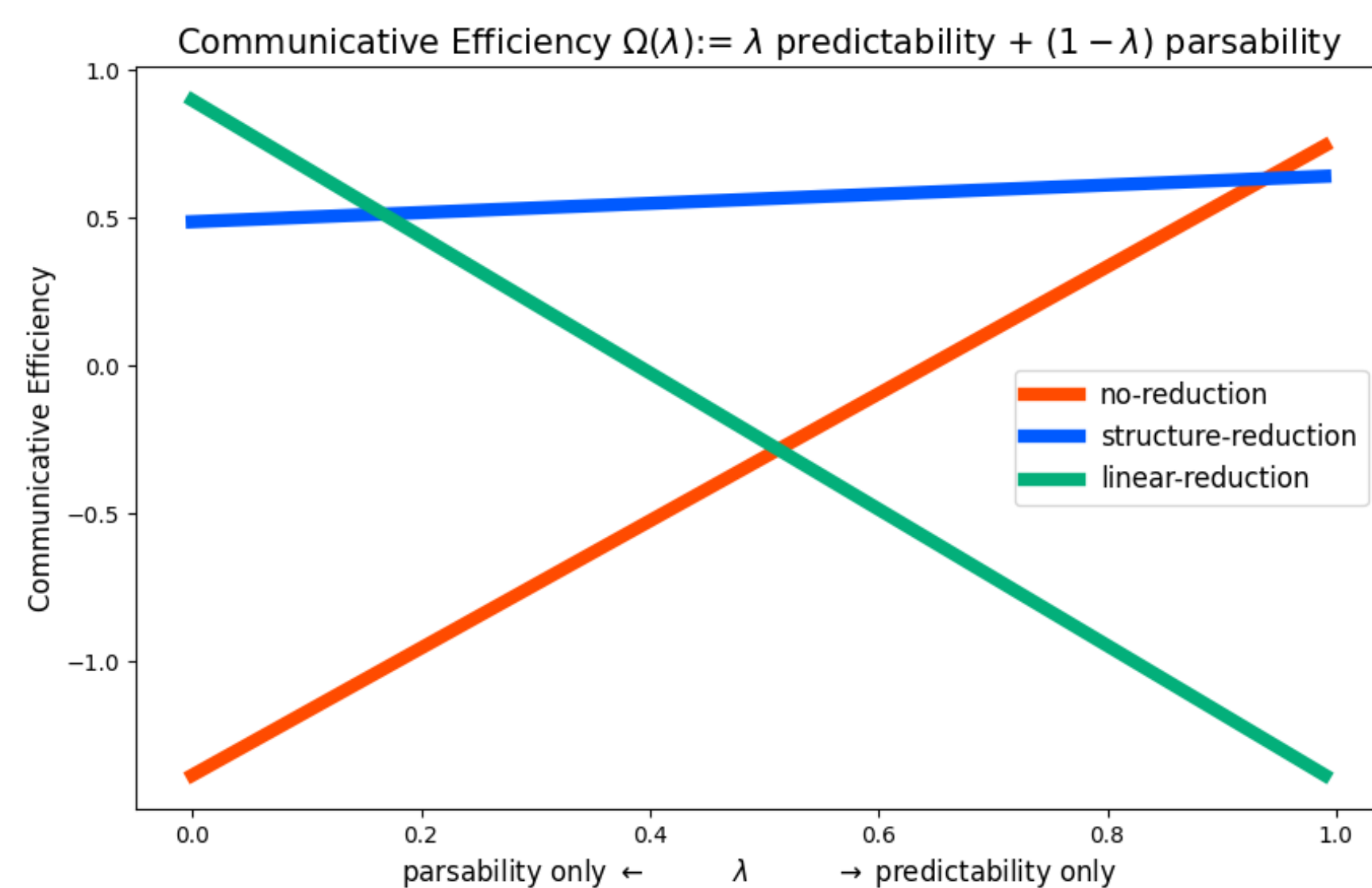
$$\text{communicative efficiency} := \lambda \cdot \text{predictability} + (1 - \lambda) \cdot \text{parsability}, \quad (\lambda \in [0, 1])$$

3種類の言語を設計

- no-reduction言語: Mary called John and Mary praised John
- structure-reduction言語: Mary called ____ and ____ praised John
- linear-reduction言語: Mary called John and ____ praised ____

	縮約あり	縮約なし
構造依存	structure-reduction	no-reduction
線形	linear-reduction	

結果：構造依存な縮約をもつ言語が（多くの場合）最適解



実験詳細

- 人工PCFGs [White & Cotterell, 2021]で64種類の語順の人工言語コーパスを作成
- それをもとに、縮約操作が異なる3種類の言語をそれぞれ作成
- predictabilityとparsabilityは、Recurrent Neural Network Grammars [Dyer et al., 2016]で推定

Case Study2: 自由語順と項省略 [梶川ら, 2025]

文法相関もまた領域一般な認知能力から説明可能か？

- 自由語順と項省略の存在には相関関係がある [Hale, 1980; Oku, 1998; Saito, 2002; Takahashi, 2008]
- こうした相関はなぜ存在するのか？領域固有のパラメータ？それとも？

自由語順（かき混ぜ）

花子が 太郎を ほめた (主語—目的語—動詞)

太郎を 花子が ほめた (主語—目的語—動詞)

項省略

花子が 太郎を ほめた (主語の省略)

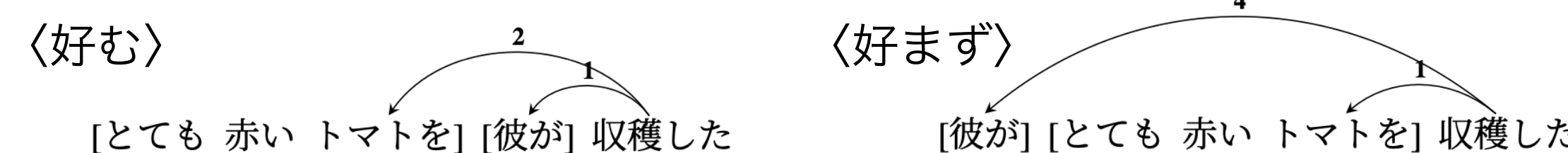
花子が 太郎を ほめた (目的語の省略)

記憶と予測のトレードオフ

- 単純性内で、記憶と予測の間にはトレードオフ関係がある
 - 心理言語学研究において、この2つが文の処理負荷の要因 [Gibson, 2000; Lewis & Vasishth, 2005; Isono, 2024; Hale, 2001; Levy, 2008; Futrell et al., 2020]

記憶に関する理論：Dependency Locality Theory (DLT) [Gibson, 2000]

- 依存関係の距離が遠いほど、作業記憶での手続きで負荷が生じる
- “long-before-short”語順 [Hawkins, 1994; Yamashita & Chang, 2001]の選好も説明可能：



予測に関する理論：Surprisal Theory [Hale, 2001; Levy, 2008]

- 単語の処理負荷は、その予測のしにくさに比例する：

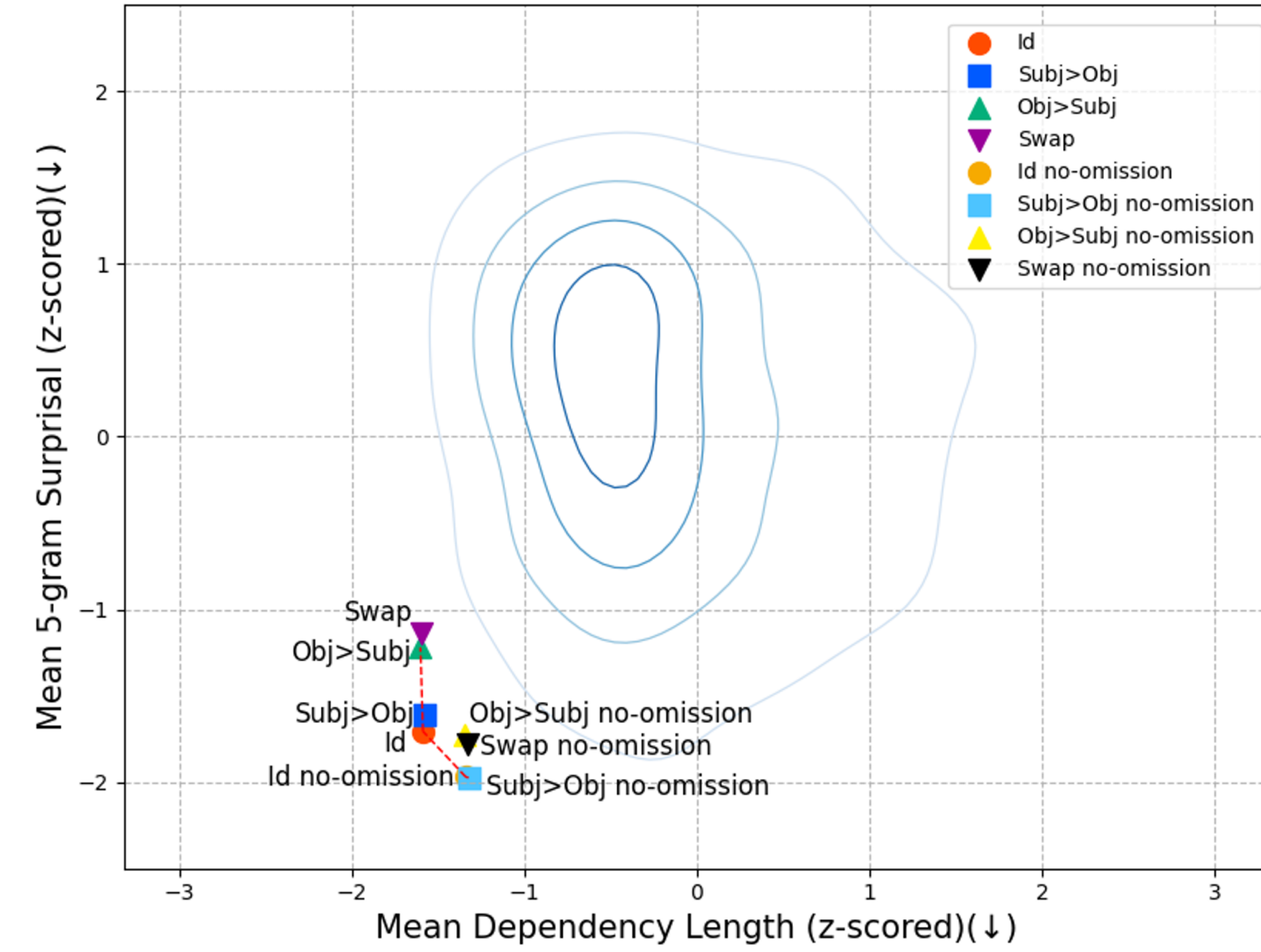
$$\text{difficulty}(w) \propto -\log P(w | \text{Context})$$

- 項省略は、後続動詞のsurprisalが大きくなるように行われている [石月ら, 2022]

- 記憶と予測にはトレードオフ関係がある [Futrell et al., 2020; Hahn et al., 2021]
- 自然言語の語順は、このトレードオフのもとで、効率的なように形作られている [Gildea & Jaeger, 2015; Hahn et al., 2021; Hahn & Xu, 2022]
- 自由語順と項省略は、記憶と予測の処理負荷に大きな影響を与える

自由語順と項省略の相関関係は、記憶と予測のトレードオフにより説明できる？

結果：自由語順と項省略両方がある日本語は最適解の1つ



$$\Omega_L(\lambda) := \lambda \cdot \text{deplen}_L + (1 - \lambda) \cdot \text{surp}_L, \quad (0 \leq \lambda \leq 1)$$

$$L^* = \arg \min_{L \in \mathcal{L}} \Omega_L(\lambda)$$

最も効率的な言語L*	λの範囲
Subj>Obj no-omission	[0.00, 0.11]
Id no-omission	[0.12, 0.52]
Id	[0.53, 0.97]
Obj>Subj	[0.98, 1.00]

LLMのような「予測偏重」型モデルが普及すると、項省略はなくなる！？

実験詳細

- UD_Japanese-BCCWJ (v2.10, 長単位分割) [Asahara et al., 2018]とJAOJ [Ishizuki et al., 2024]を用いて、
 - 依存構造と、
 - どこでどのような項が省略されているかの情報を獲得
- 3,428文、67,088単語+2,895単語分の項が省略
- Id: 語順をそのまま、Subj>Obj: 語順をS—Oに固定、Obj>Subj: 語順をO—Sに固定、Swap: S—OとO—Sの関係の逆転、に加え、項省略あり/なしの**仮想日本語**コーパスを作成
- ベースラインとして、語順規則をランダムに再構築した文法×500を設計し、項省略あり/なしの合わせて1,000の言語を作成