

Inflecting When There's No Majority: Limitations of Encoder-Decoder Neural Networks as Cognitive Models for German Plurals

Kate McCurdy, Sharon Goldwater, Adam Lopez (Edinburgh)

ACL 2020

能地宏
産総研

Kind → Kinder

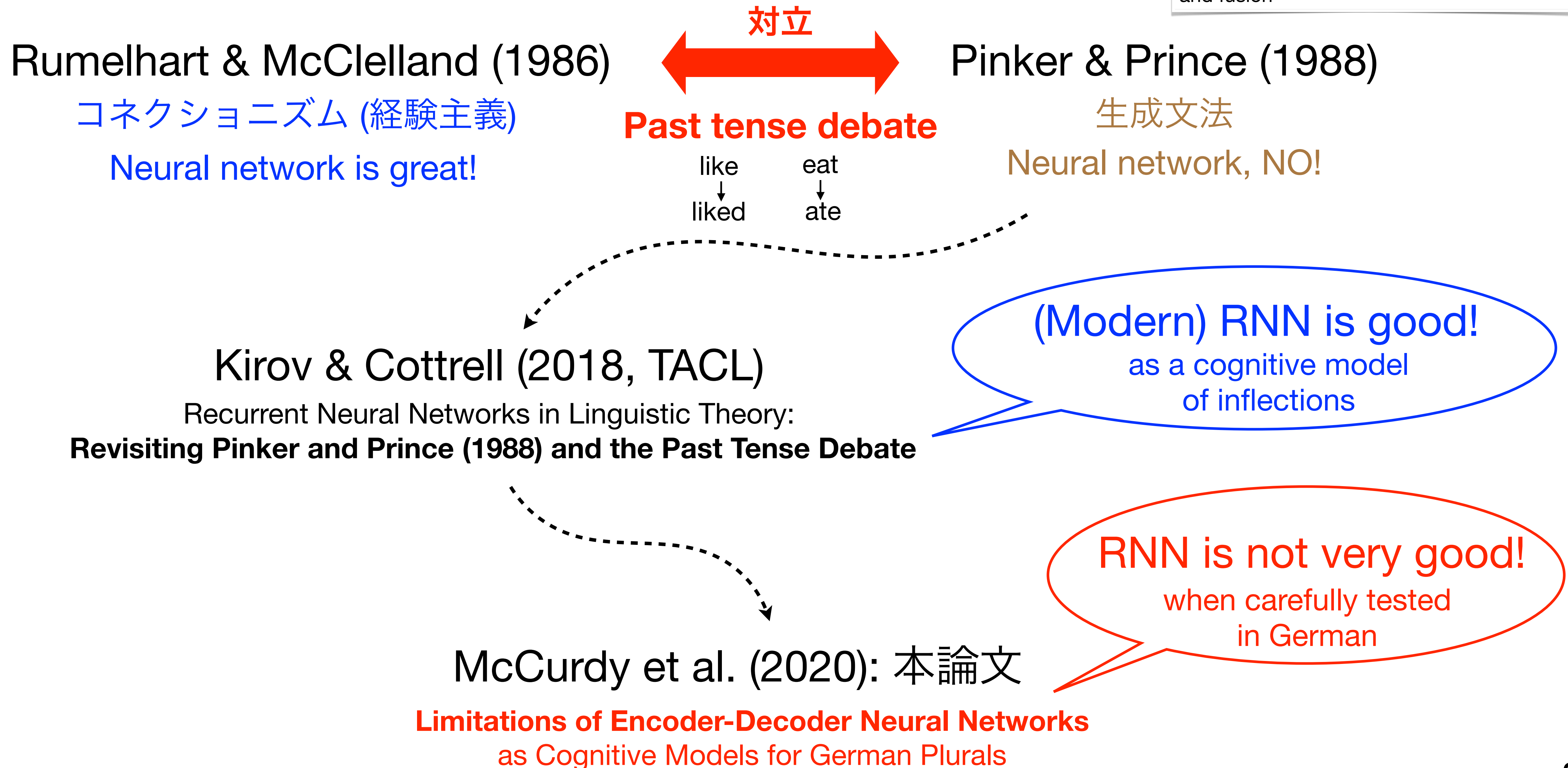
Bier → Biere

Straße → Straßen

論文の位置づけ

おすすめ論文 (中立な立場?)

Pater (2018): Generative linguistics and neural networks at 60: foundation, friction, and fusion



以下の3つに焦点を当てて説明

- ニューラルネットによる認知モデルとは？主にそのモチベーション
- Past tense debate と Kirov and Cottrell (2018) の主張
- 本論文で主張している RNN（現在のニューラルモデル）の問題点
（恐らく）工学的応用とも関連

いわゆる計算心理言語学

- 目的：数理モデルの構築・比較による、人の言語処理の仕組みの理解
- *ACL 系で同じカテゴリに属すると思われる研究
 - RNN による agreement error と人の統語処理の比較（篠田さんの発表）
 - RQ: RNN (LSTM) と人の振る舞いがどれだけ似ているか？（人の文処理は LSTM 的か？）
- （恐らく）*ACL 系で受け入れられてる理由/背景
 - Interpretability/black box NLP への貢献と見なせるから
 - ニューラルが何ができて何ができないか（工学への貢献）

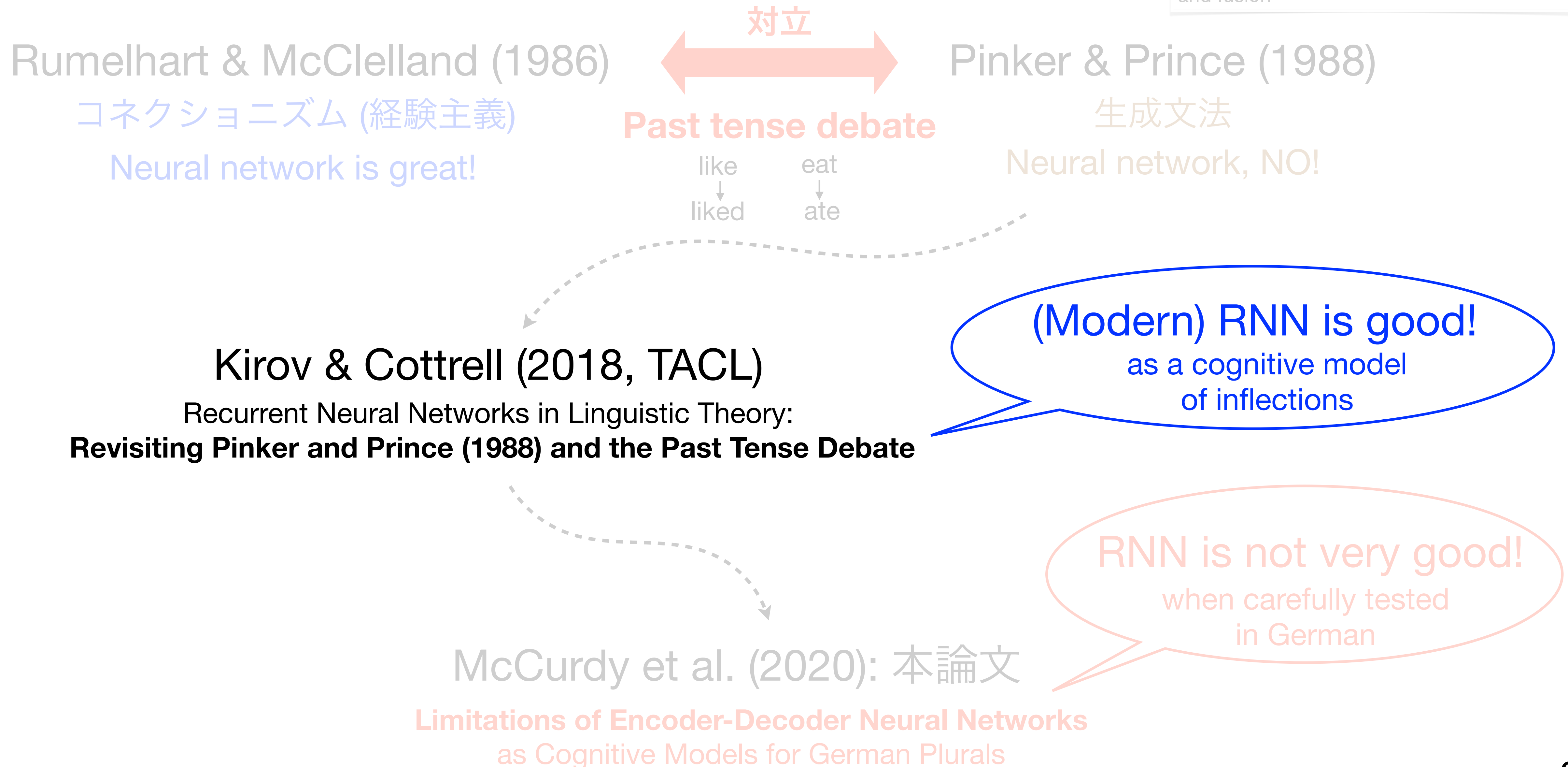
本研究の焦点: 単語の活用

- 人は単語の活用をどのように処理しているか？ (like→liked; swim→swam)
- Past tense debate (コネクショニズム vs. 生成文法)
 - 一つの論点:
規則変化と不規則変化は脳が別々に処理しているか？ 単一モジュールか？
- Rumelhart & McClelland: 単一のニューラルネットで全て説明できる
- Pinker & Prince: それぞれ別のモジュールで別の“規則”によって処理される
(当時の) ニューラルネットはあれもできないこれもできない、精度が低すぎる、etc.と批判
- cf.) Top-down vs. bottom-up (菅原さんの話と関連)

論文の位置づけ

おすすめ論文 (中立な立場?)

Pater (2018): Generative linguistics and neural networks at 60: foundation, friction, and fusion



Kirov & Cottrell が示したこと

orthographic		IPA		infl. type
stem	past	stem	past	
go	went	gou	went	suppletive
sing	sang	siŋ	sæŋ	ablaut
swim	swam	swim	swæm	ablaut
sack	sacked	sæk	sækt	[-t]
sag	sagged	sæg	sægd	[-d]
pat	patted	pæt	pætid	[-id]
pad	padded	pæd	pædid	[-id]

gou → went
sæk → sækt
...

Encoder-decoder
(LSTM+attention)

wiðdró: → ?
(withdraw)
siŋ → ?
(sling)

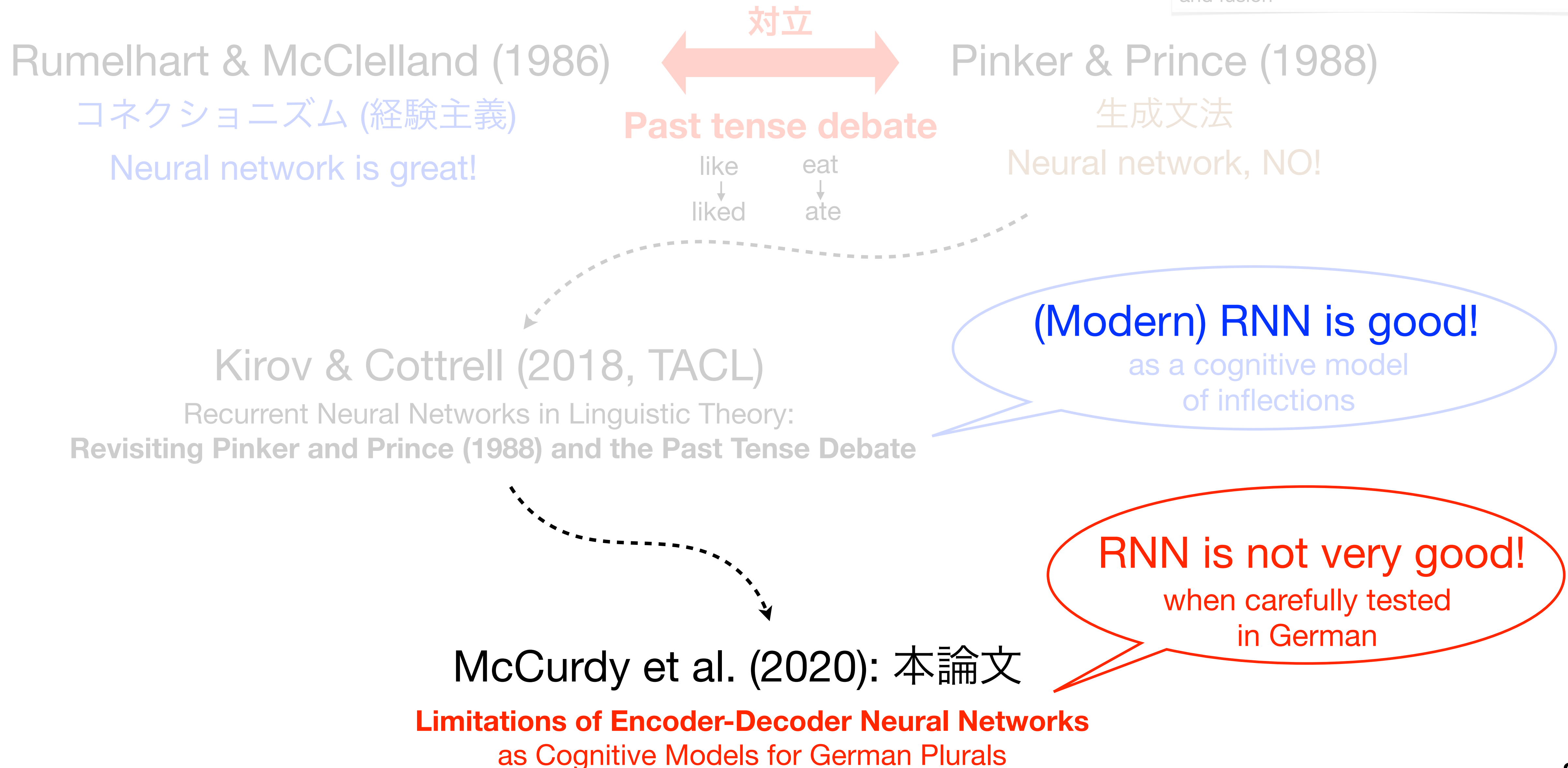
現在形音素列 → 過去形音素列の変換の教師データ
~3200単語

- 彼らの Enc-dec の精度は R&M (1986) の推定精度 (~67%) を遥かに上回る (~97%)
- 未知語に対して人と同じエラーの傾向を示す
 - 規則変化への間違いが多く、逆はほぼ起きない (throw → *throwed) ⇒ 認知的には妥当と主張

論文の位置づけ

おすすめ論文 (中立な立場?)

Pater (2018): Generative linguistics and neural networks at 60: foundation, friction, and fusion



本論文 (McCarthy et al. 2020) の主張

- K&C 2018 の Enc-dec がうまくいったのは英語の活用が比較的単純だからであり、人間の活用能力一般をモデル化できていない、と主張
- 特に問題として挙げている点
 - エラー傾向 (throw→threw) の原因は、
**Enc-dec が人の処理と似ているからではなく、
英語では規則変化が1種類 (-ed) で、それが最も高頻度だからではないか**
(音素で -t, -id などの変化はあるがニューラルはそれらは学習できるだろう)
 - 人は頻度に引きずられない一般化を行う場合がある (e.g., ドイツ語の活用)
 - ドイツ語を活用を学習した Enc-dec は、人と同じエラー傾向を示すか？ ⇒ NO

ドイツ語の名詞の活用 (大きく5種類)

Suffix	Singular	Plural	Type	Token
/-(e)n/	Strasse	Strassen	48%	45%
/-e/	Hund	Hunde	27%	21%
	Kuh	Kühe		
/-Ø/	Daumen Mutter	Daumen Mütter	17%	29%
/-er/	Kind	Kinder	4%	3%
	Wald	Wälder		
/-s/	Auto	Autos	4%	2%

女性名詞に多い

Der Strasse → Die Strassen

男性/中性名詞に多い

Den Hund → Die Hunde

男性/中性名詞に多い

外来語に多い

Das Autos → Die Autos

Table 1: German plural system with examples, ordered by CELEX type frequency (Sonnenstuhl and Huth, 2002).

- 英語話者は新語を規則変化しやすい (wag-test; wag->wagged) ことが知られる
- ドイツ語話者が新語をどのように活用するかは未知な部分が多い
 - /-s/ は低頻度だが外来語に使えるため、新語にデフォルトとなる説 (Marcus et al., 1995)
 - それよりも性の影響 (冠詞 der, den, das) が強いという説 (Zaretsky & Lange, 2016)

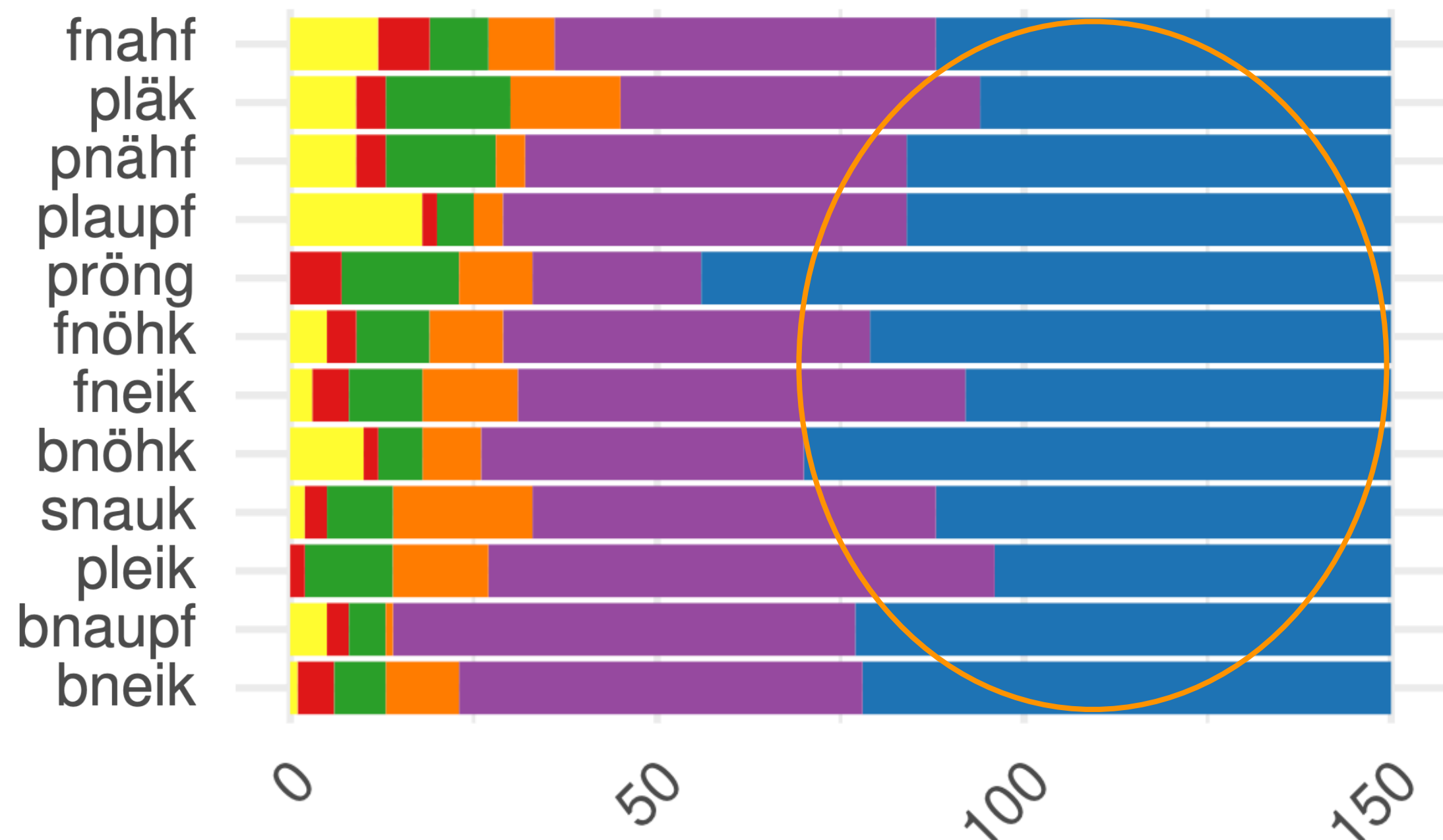
新たに母語話者の反応のデータ収集

- 過去収集したデータにはそれぞれ問題があるため、本研究で新たに母語話者の反応を収集
- 各新語 (ドイツ語っぽい単語) を150人の母語話者に提示
- 性の影響をなくするため全て中性名詞と仮定
 - Das Bral → Die _____

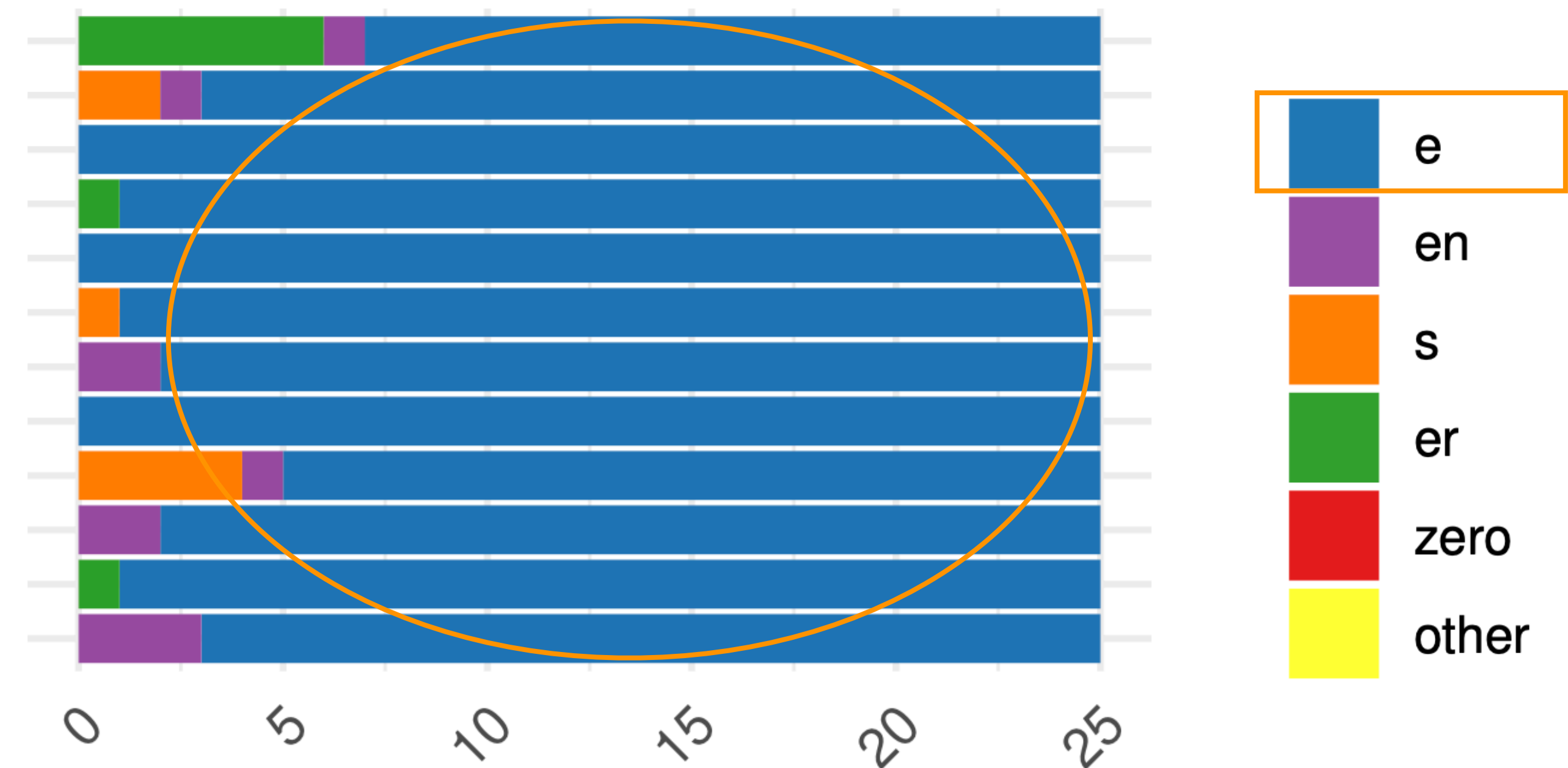
Rhymes	Non-rhymes
Bral	Bnaupf
Kach	Bneik
Klot	Bnöhk
Mur	Fnahm
Nuhl	Fneik
Pind	Fnöhk
Pisch	Plaupf
Pund	Pleik
Raun	Pläk
Spand	Pnähf
Spert	Pröng
Vag	Snauk

Marcus et al. (1995) で
使われた 24 の新単語

人の反応の分布と、モデルの生成結果の分布



各語に対する150人の反応の分布



ドイツ語活用データで Enc-dec モデルを
複数学習 (25 random seeds) → 予測結果を集計

- モデルは単語によらず /-e/ と活用してしまう (~90%)
 - 訓練データの中性名詞の高頻度は /-e/ ⇒ 高頻度をデフォルトにしてしまうのでは？
- 人間も約半数は /-e/ だが、 /-en/ と予測する事例も多い (~34%)

腑に落ちない点

- 異なる random seed のモデルの結果を集計することの妥当性？
 - 訓練データ/ハイパーパラメータは共通のため、(特に top-1 が) 似た予測を行うモデルが生じるのは当たり前に見える
 - 著者らは、英語でシングルモデルの結果の分散が大きいことが分かったため、と主張している
 - 別の実験で 2-best, 3-best の結果に /-en/ が出現するかも調べているが、むしろ変な語 (others) が増えるという結果に
cross-entropy だから当たり前、とも言えるが、人間は学習時に正解の活用以外目にしない

結論

- Enc-dec の学習するデフォルトの振る舞いは、頻度の影響を大きく受けてしまう（人はそうとは限らない）

(Non-Rhymes). Regardless of the minority-default question, it seems that ED models do not necessarily function as good cognitive approximations for inflectional systems like German number, in which no class holds the majority.

所感

- 実験結果は必ずしも綺麗な結果ではないが、 /-s/ がデフォルトとは言えないことを定量的に調べた点なども高く評価されたと推測
- 著者達自身は必ずしもニューラルの枠組み自体を否定しているわけではないのでは
 - 現在のモデル/学習法でうまく扱えない言語現象（限界）を示した
⇒ それを補うモデル (inductive bias) を考える、という流れ？
- しかし真理を求めるのがゴールであると考えたと、この方向で良いのかは検討が必要
 - 正しい丘に登っているのか？ (Top-down vs. bottom-up; Bender & Koller, 2020)