

# A Computational Linguistic Study of the Omission of Third-Person Singular -s in English

岡村康平

## 1. はじめに

- (1) 第二言語 (L2) の十分なインプットを受けた L2 学習者 (L2er) にも見られる、一貫性のない L2 使用 (可変性)。  
(Haznedar & Schwartz (1997))
- (2) 可変性の要因を巡る、L2 言語習得 (SLA) 研究上の仮説
  - a. 中間言語の欠陥を主張 (臨界期以後、L1 にない解釈不可能素性にはアクセス不可能):  
Representational Deficit Hypothesis (RDH) (Hawkins and Liszka (2003))
  - b. 中間言語外の要因を主張 (臨界期以後も解釈不可能素性にアクセス可能)  
Missing Surface Inflection Hypothesis (MSIH) (Haznedar & Schwartz (1997), Prévost & White (2000))
- (3) 各文法項目の研究で MSIH が支持されてきている背景
  - a. L2er の当該文法項目の使用規則に一貫性が見られること
  - b. L2er の当該文法項目の使用に、NS と共通した性質が見られること
- (4) 文法誤り影響を及ぼすとされる 3 要因
  - a. 一般認知領域での処理
  - b. 派生時の計算処理
  - c. 中間言語の欠陥
- (5) 現在 MSIH の検証に用いられる主な実験方法
  - a. 文法性判断課題
  - b. 自己ペース読み課題 (各単語に対する反応時間の計測)
- (6) しかし、英語学習者による 3 人称単数現在 (3 単現) の -s の脱落については、(3) に挙げられた手法を用いた研究が実施されているが、MSIH を支持する証拠が存在しない (cf., Wakabayashi et al. (2018))
- (7) 3 単現習得の研究が困難を極めている理由
  - a. 主述一致という統語論と形態論を包括する現象である
  - b. 厳密に形態素の選択規則を規定した言語理論の少なさ
  - c. 素性の列挙から形態素の選択までの処理を区別して行う実験方式がないこと  
→ 自己ペース読み課題の限界
  - d. 理解しているにも関わらず犯してしまう稀有なケースの文法誤りであること  
(Wakabayashi et al. (2018))
- (8) Frimu and Dekydtspotter (F&D) (2022) による、Distributed Morphology (DM) 基盤の強制ペース読み課題
  - これまで実験では得られなかった微量な反応を感知可能な画期的実験方式
  - ⇒ 言語領域・一般認知領域のインターフェース上での反応時間を計測することで、NS と L2er に共通するフランス語使用上の一貫性を示唆

- (9) Okamura, Kimura and Kato (2022) は、強制ペース音読課題を通して、英語の 3 単現の-s において実験を行ったところ、英語はフランス語と変化形の数や、語彙の指定が大きく異なるにも関わらず、L2 学習者から同様の結果が得られた (実験 1)。  
→ 中間言語の要因を排除できる
- (10) Okamura, Kimura and Kato (2022) を踏まえた 3 単現脱落の要因  
a. 一般認知領域での処理  
b. 派生時の計算処理  
e. 中間言語の欠陥
- (11) しかし、認知要因と計算要因を切り分けることは、現状の心理言語学的アプローチでは不可能  
→ そこで、大規模言語モデル (Large Language Model: LLM) (特に Syntactic Language Model) を用いた実験を行い、計算要因に脱落の要因があることを示す  
⇒ LLM には一般認知領域が存在しないため、L2er と同様の傾向が見られた場合、言語の構造に脱落の要因が存在することの証左となる
- (12) The Anti-Lexicalist Theories (cf. Distributed Morphology (Halle and Marantz (1993)))
- (13) The Single Engine Hypothesis:  
There is only one generative engine to build complex words and sentences.
- (14) Hierarchical Morphological Processing Hypothesis (Oseki (2018: 58)):  
“Morphological processing tracks hierarchical syntactic structures of words.”
- (15) どちらのモデルがより人間の処理により近いのか(実験 2):  
a. 形態素を末端節点としてトレーニングした Syntactic Language Model  
b. 語を末端節点としてトレーニングした Syntactic Language Model  
→ 形態素を末端節点とした LLM は、高い言語能力が得られ、人間の言語処理機構を仮定するモデルとして妥当であることが示された  
⇒ Hierarchical Morphological Processing Hypothesis の支持
- (16) 3 単現脱落の原因が英語の統語的形態論的性質に依存するのか(実験 3)  
Penn Treebank (PTB) 形式コーパスで訓練された LLM を用いて、英語学習者による 3 単現の脱落に関わる現象が、一般認知領域の要因を排除した環境でも観察されるかを検証する

**RQ** 3 単元の-s 脱落の要因は、中間言語の欠陥ではなく、英語の構造に起因するか

論文構成：

2. 心理言語学的アプローチ
  - 2.1 3 単現の-s 習得の課題
  - 2.2 実験 1 の概要
  - 2.3 結果
3. 計算言語学的アプローチ

- 3.1 Liner Language Model と Syntactic Language model
- 3.2 実験 2
  - 3.2.1 仮説
  - 3.2.2 結果と考察
- 3.3 実験 3
  - 3.3.1 実験 3.1
  - 3.3.2 実験 3.2 (現在進行中)
- 4. まとめ

## 2. 心理言語学的アプローチ

### 2.1 Okamura, Kimura and Kato (2022)

- (17) a. Les enfants mangeront. [Agreement singular condition]  
 The children eat-FUT.3PS.PL
- b. L'enfant mangera. [Agreement plural condition]  
 The child eat-FUT.3PS.SG
- c. \*Les enfants mangera. [Error singular condition]  
 The children eat-FUT.3PS.SG
- d. \*L'enfant mangeront. [Error Plural condition]  
 The child eat-FUT.3PS.PL  
 'The child/ the children will eat.'

(Frimu and Dekydtspotter (2022:6-7), with modifications)

### (18) Activated vocabulary items in DM

Syntactic structure contains:	Activations of subset vocabulary items in DM
[T: <i>u</i> Person: Ø, <i>u</i> Number: Ø] (3PS.SG agreement features)	a) /-a/ ↔ [T: <i>u</i> Person: Ø, <i>u</i> Number: Ø]
[T: <i>u</i> Person: Ø, <i>u</i> Number: Plural] (3PS.PL agreement features)	a) /-a/ ↔ [T: <i>u</i> Person: Ø, <i>u</i> Number: Ø] b) /ɔ̃T/ ↔ [T: <i>u</i> Person: Ø, <i>u</i> Number: Plural]

(Frimu and Dekydtspotter (2022:5), with modifications)

### (19) Vocabulary of English (fragment)

- a. [3SG, PRES] ↔ -s
- b. [PRES] ↔ -Ø (Bobaljik (2017:6))

### (20) a. -s ↔ [*u*Person: 3p, *u*Number: Singular]

- b. -Ø ↔ [*u*Person: Ø, *u*Number: Ø] (Okamura, Kimura and Kato (2022))

(21) Activated vocabulary items in different condition in number

	Representation in Syntax:	Vocabulary items:
3PS.SG: The child eats a lot of desserts.	[ <i>u</i> Person: 3p, <i>u</i> Number: Singular]	(20a), (20b)
3PS.PL: The children eat a lot of desserts.	[ <i>u</i> Person: 3p, <i>u</i> Number: Ø]	(20b)

- (22) a. The child eats a lot of desserts. [Agreement Singular condition]  
 b. The children eat a lot of desserts. [Agreement Plural condition]  
 c. \*The child eat a lot of desserts. [Error Plural condition]  
 d. \*The children eats a lot of desserts. [Error Singular condition]
- (Okamura, Kimura and Kato (2022))

(23) [The child] [eats] [PIC] [a lot of] [desserts.]

(24) Details of participants

Proficiency	<i>n</i>	Average age	Estimated proficiency
Beginner	20	19.2 (18-20)	CEFR A1-A2
Intermediate	9	23.9 (21-47)	CEFR B1-B2
Advanced	3	22.3 (21-23)	CEFR C1
Total	32		

(Okamura, Kimura and Kato (2022))

● 結果

(25) 正しい形態素に対する便感度は、単数形の方が高い

<i>d'</i> singular verb form	<i>d'</i> plural verb form
1.062 (0.90)	0.500 (0.90)

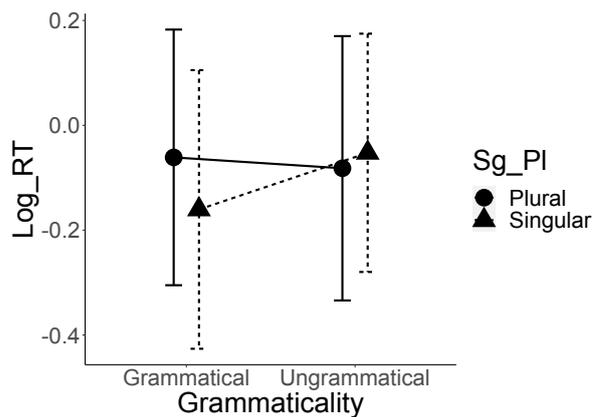
(Okamura, Kimura and Kato (2022))

(26) 画像分類課題における反応時間 (RT)

Error Singular	Agreement Singular	Error Plural	Agreement Plural
901 (194)	825 (197)	875 (217)	865 (210)

(Okamura, Kimura and Kato (2022))

(27) The result of GLM (General Linier Model)



(Okamura, Kimura and Kato (2022))

(28) 結果のまとめ

- a. 現在形の一致における  $d'$  の値が 0 ではないことは、L2er は形態素を正しく選択できる能力があることを示している。  
→ これまでの日本人英語学習者に関する先行研究とは異なり、MSIH を支持する結果
- b. 画像選択課題における RT と、主語名詞句の単複の違いとの間に有意差があったことは、意味解釈処理と形態素選択に関連する処理が同じ処理領域で競合していることを示す  
→ 形態素選択を補助する推論 (一般認知領域) が存在する可能性を主張する Frimu and Dekydtspotter (2022) を支持する結果

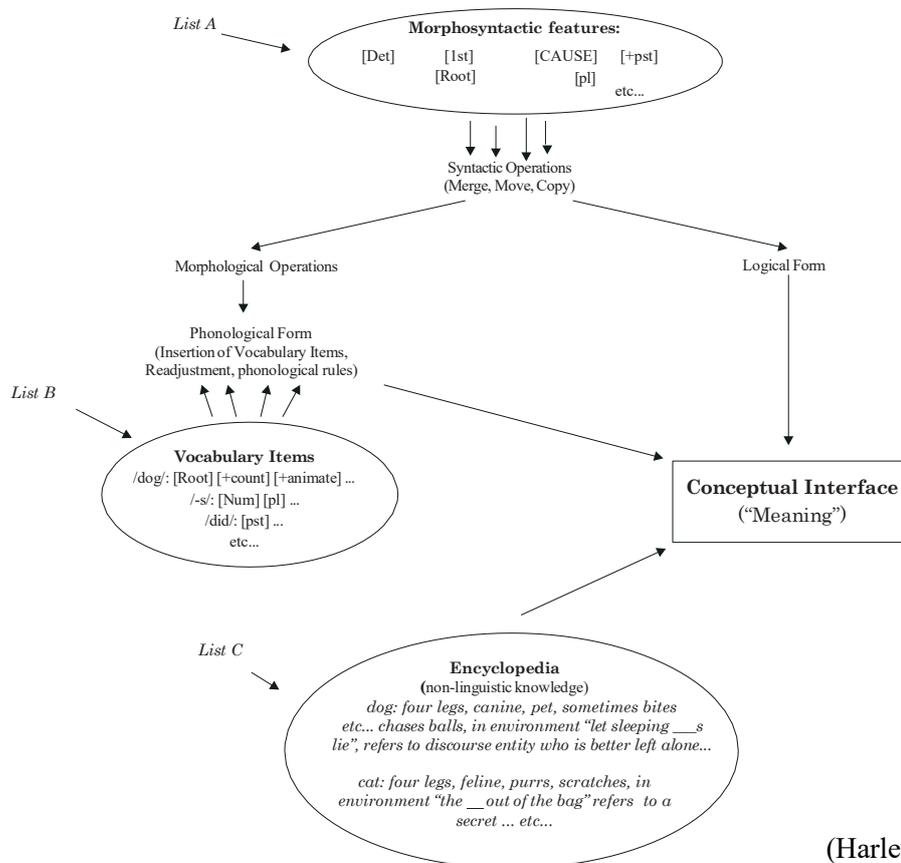
● 考察

(29) 実験 1 で明らかになったこと

- a. 形態素選択に関連する処理 (Frimu and Dekydtspotter は形態素選択を補助する推論と説明) と意味処理の競合が存在する
- b. a の競合は L2er の間で一貫して見られた  
→ MSIH が支持された

2.2 中間取りまとめ (Interim Conclusion): 3 単現脱落の要因のありか

(30) 文法アーキテクチャ



(Harley and Noyer (1999:3))

(31)

	計算要因あり	計算要因なし
一般認知要因あり	○	○
一般認知要因なし	○	×

(32) 今後の課題

- a. MSIH を支持する結果が得られたものの、依然として L2er の 3 単現の -s 脱落の要因は明らかではない
  - b. 心理言語学的アプローチでは、現在一般認知領域での処理と計算上の処理を切り離れた実験方式は存在しない
- この問題を解決するため、大規模言語モデルによる実験を行う
- ⇒ 人間のような一般認知領域が存在しない大規模言語モデルの処理で、心理言語学的アプローチと同様の結果が得られた場合、計算要因に 3 単現の -s 脱落の要因が存在する可能性が示唆される

### 3. 計算言語学的アプローチ

#### 3.1 扱う LLM とその訓練方式

- LLM の選択

(33) Liner Language Model

LLM で主流な transformer モデル (Vaswani (2017)) は、 $n$ -gram をマルコフ過程に適用する統計的手法を用いる (feed-forward networks)。

e.g. GPT-3 (generative pretrained transformers) (Brown (2020)), LLaMA (language model for dialogue applications) (Touvron (2023)), PaLM2 (pathways language model, version 2) (Anil (2023))

→  $n$ -gram ( $n$ =入力トークン数) を使用した線形順序に基づくモデル

$$P(x_1, x_2, \dots, x_T) = \prod_{t=1}^T P(x_{<t}) \quad (2)$$

- 各トークンの確率算出方式 (Multi-Head Self-Attention による確率予測<sup>1</sup>)

$$Attention(Q, K, V) = softmax\left(\frac{(QK^T)}{\sqrt{d_k}}\right)V \quad (3)$$

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (4)$$

where  $head_i = Attention$

(34) Syntactic Language Model

モデルごとに異なる処理を用いて統語構造を機械学習の対象とする。本研究ではその中でも代表的な Recurrent Neural Network Grammar (RNNG) (Dyer et al. (2016)) を扱う。

(35) RNNG は  $n$ -gram を基盤としているが、構文解析結果 ( $y$ ) を組み込む点で Transformer と異なる<sup>2</sup>。さらに、RNN に LSTM (Long Short Term Memory) を組み込むことで長距離依存に対処している。

a. 訓練データ : Penn Treebank (PTB) 形式データ (もしくは通常の文データ)。

b. 訓練方式 : 教師あり学習<sup>3</sup>

c. モデル : Noji and Oseki (2021) による python ベースの RNNG<sup>4</sup>

---

<sup>1</sup>本文中で説明のされていない self-の由来は、Query, Key, Value の初期値が  $Q=K=V$  であることによる。これらの値はランダムに初期化される。

<sup>2</sup> $r$  はアクション (S (Stack), T (Buffer)) に対する重みベクトル、 $b$  はバイアスベクトル、 $n$  は非ターミナル (e.g. S, NP) の埋め込みベクトル、 $u_t = \tanh(W[o_t; s_t; h_{(t)}] + c)$  ( $o, s$  は T, S の出力、 $H$  はアクション履歴 ( $a_{<t}$ )、 $W$  はパラメーター、 $c$  はバイアスベクトル)

<sup>3</sup>教師なし学習用のモデル (Kim et al (2019)) も存在するが、今回の研究の趣旨を踏まえ、教師なしモデルを用いる

<sup>4</sup>今回このモデルを使用する根拠は、初期の RNNG (Dyer et al. (2016)) よりも処理の効率化が実現されているため

(36) 構文解析の例

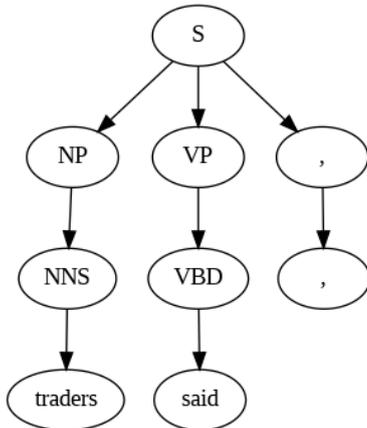
Input: The hungry cat meows .

	Stack	Buffer	Action
0	The   hungry   cat   meows   .	NT(S)	
1	(S	The   hungry   cat   meows   .	NT(NP)
2	(S   (NP	The   hungry   cat   meows   .	SHIFT
3	(S   (NP   The	hungry   cat   meows   .	SHIFT
4	(S   (NP   The   hungry	cat   meows   .	SHIFT
5	(S   (NP   The   hungry   cat	meows   .	REDUCE
6	(S   (NP The hungry cat)	meows   .	NT(VP)
7	(S   (NP The hungry cat)   (VP	meows   .	SHIFT
8	(S   (NP The hungry cat)   (VP meows	.	REDUCE
9	(S   (NP The hungry cat)   (VP meows)	.	SHIFT
10	(S   (NP The hungry cat)   (VP meows)   .		REDUCE
11	(S (NP The hungry cat) (VP meows) .)		

$$P(x, y) = \prod_{t=1}^{|a(x,y)|} P(a_{<t}) \quad (5)$$

$$= \prod_{t=1}^{|a(x,y)|} \frac{\exp \exp r_{a_t}^T u_t + b_{a_t}}{\sum_{a' \in A_G(T_t, S_t, n_t)} \exp \exp r_{a'}^T u_t + b_{a'}}$$

(37) (S (NP (NNS traders)) (VP (VBD said)) (, ,))



- 3 単現の実験を行ううえでの形態論の取り扱い

(38) The Anti-Lexicalist Theories (cf., Distributed Morphology (Halle and Marantz (1993))) (=12))

(39) The Single Engine Hypothesis:  
There is only one generative engine to build complex words and sentences. (=13))

(40) Hierarchical Morphological Processing Hypothesis (HMPH) (Oseki (2018)):  
Morphological processing tracks hierarchical syntactic structures of words. (=14))

(41) Oseki (2018) は人間の言語処理の性質に基づき、自然言語処理モデルも同様に (13) を基盤にする必要があるとしている (Oseki (2018: 168))

→ 語の多義性 (e.g., unlockable)

$$P(m_1, \dots, m_k) = \sum_{T: \text{yield}(T)=m_1, \dots, m_k}^n P(t_i) \quad (1)$$

(42) 本節では、統語構造に基づいて処理を行う Syntactic Language Model を形態論にも適用して実験を行う。

→ ただし、Syntactic Language Model 構文解析の処理をそのまま形態素解析に適用する自然言語処理の研究は未発達である

→ 3 単現の実験を行う前に、morphosyntax を仮定した処理の妥当性を示す実験をまず行う (実験 2)

(43) 実験 2 の意義

a. 形態論と統語論を区別しない訓練方式は、自然言語を適切に処理できるのかという側面からの HMPH の検証

b. a の成立により、今回扱うモデルが人間の言語処理のモデリングとして妥当であることが示される

→ 人間の生成する言語の構造を正確に解析できる訓練方式 = 人間の言語の構造を正確に学習したモデル

(44) Morphosyntax を用いて訓練した RNNG

a. 訓練データ : PTB を形態素基盤で変形したもの<sup>5</sup>

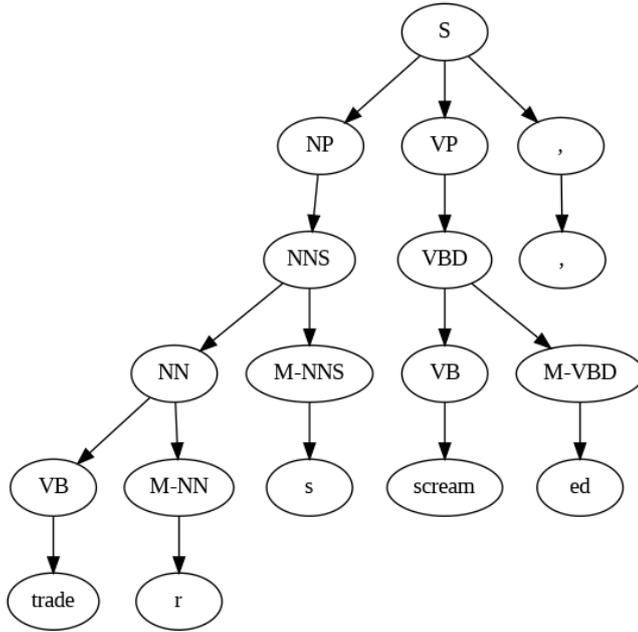
b. 訓練方式 : 教師あり学習

c. モデル : Noji and Oseki (2021) による python ベースの RNNG<sup>6</sup>

<sup>5</sup> 単語の分節化においては、Natural Language Tool Kit package (NLTK) (Bird et al. (2009)) in Python, WordNet (Miller (1995)) を使用し訓練した Morphessor (Creutz and Lagus (2007)) を使用し、単語を形態素に分けたものを一部修正した。

<sup>6</sup> 今回の実装にあたり、ある程度の変更を加えており、特に性能に変化を与えない領域において、形態素を使用することによる情報量の増加を考慮して、構文解析における Recursion の深さの最大値を約 6 倍に増やしている。ただし、これは扱える構文の階層の深さを規定する部分のコードであり、GPU の負担が増える一方で性能には何ら影響を与えない。

(45) (S (NP (NNS (NN (VB trade) (M-NN r)) (M-NNS s))) (VP (VBD (VB scream) M-VBD)) (, ,))



- もし、HMPH が正しければ、Morphosyntactic Language Model は、RNNG よりも自然な言語の処理が可能なが予測される。

## 3.2 実験 2

### 3.2.1 実験の概要と仮説

(46) 実験の概要：実験 3 を行う前に、形態素をトークンとすることの妥当性を確かめる  
 → 単語をトークンとするモデルよりも、自然な処理が得られるのであれば、人間の処理機構により近いことの証左となる

(47) 言語処理の自然さを示す指標：Perplexity (PPL)

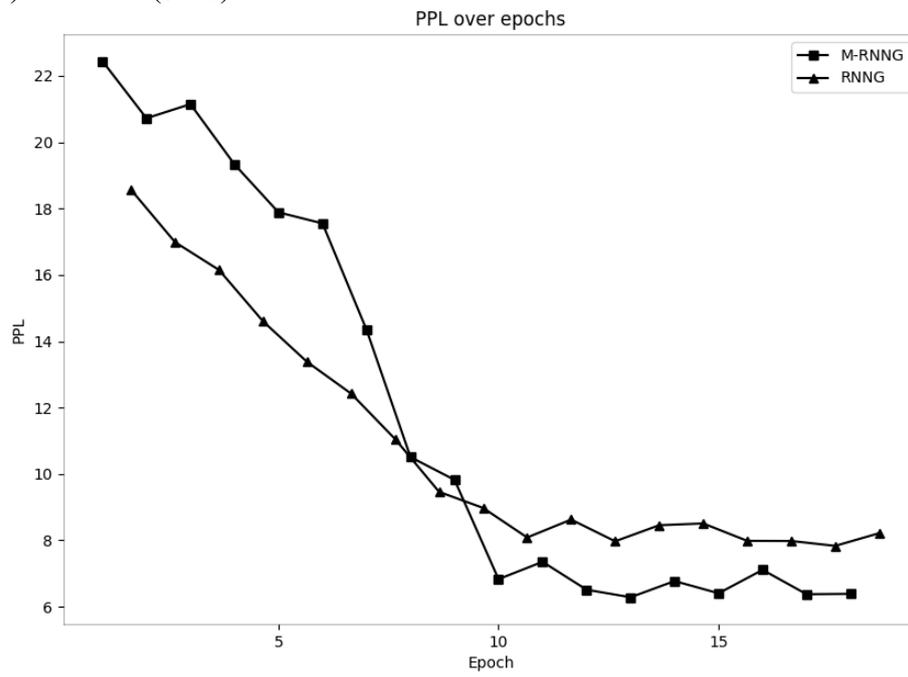
$$PPL = \exp \exp \left( \frac{total\_loss}{total\_num\_words} \right)$$

(48) 予測

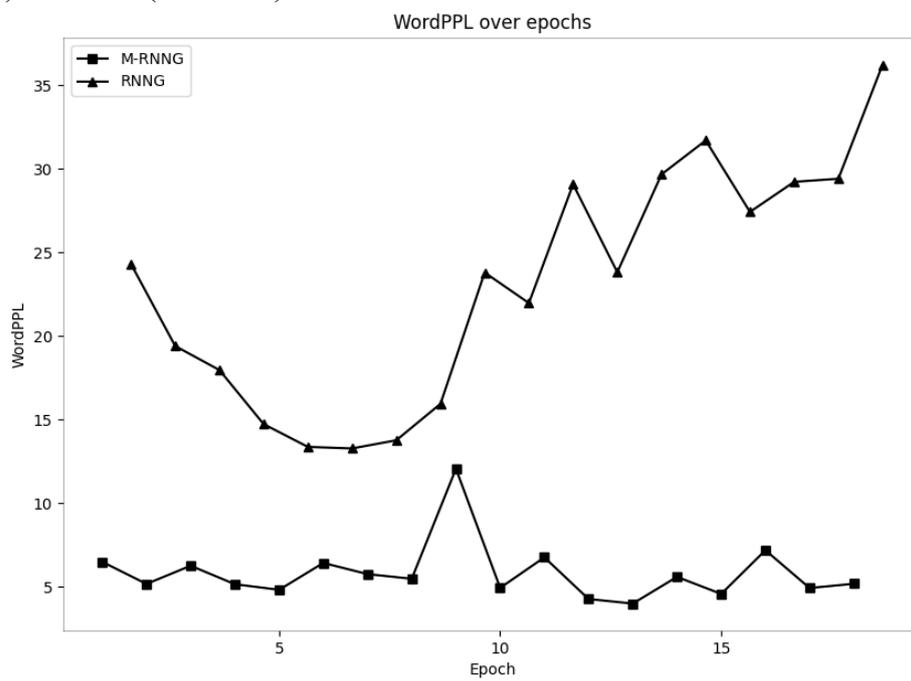
- トレーニング量が閾値を超えると、PPL が低下する。
- 少量の学習データでも低い PPL を維持することができる

### 3.2.2 結果と考察

#### (49) a. PPL (総合)



#### (50) a. PPL (トークン)



- (51) a. Morphosyntactic Language Model はデータ量やエポック数への依存が少ない
- b. WordPPL においては過学習の傾向が観察されなかった
  - Syntactic Language Model は単語同士の文法関係ではなく、単語同士の組み合わせに対して過学習を行った可能性。一方、単語同士の関係性を語根とそれ以外の形態素で分けた Morphosyntactic Language Model は、過学習による性能低下が起こりにくかったか。
- c. 閾値を超えた Morphosyntactic Language Model は、ActionPPL と総合的 PPL において、急激な低下傾向がみられた

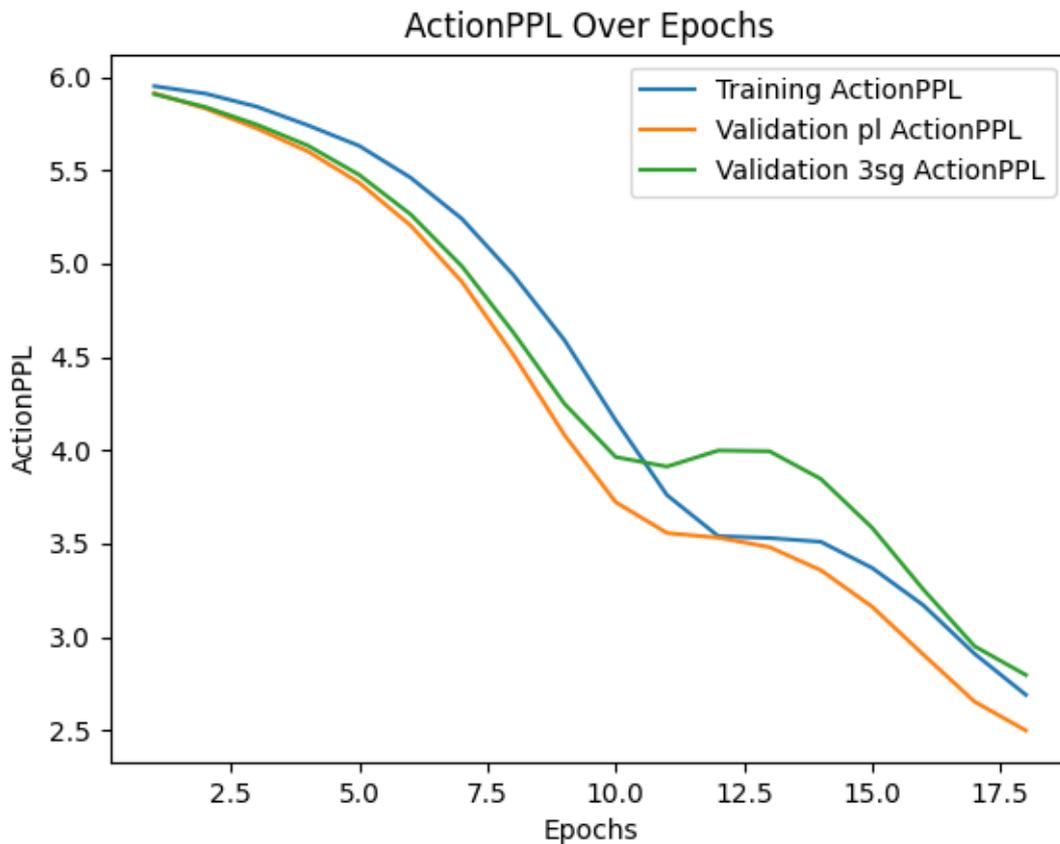
### 3.3 実験 3 (現在進行中)

#### 3.3.1 実験 3.1

- (52) 3 単現-s の誤りは認知・言語のどちらに原因があるのかを判別する
  - a. Okamura, Kato, and Kimura (2022) では、認知・言語にその原因がある可能性が示されたが、主にどちらなのかについては未検証。
  - b. 機械学習において、単数形・複数形の一致において、差が見られれば、言語にその要因がある可能性が示される。
  
- (53) 数の一致を困難にしようする要因：
  - a. 3 単現の-s が付加された動詞変化形と、それ以外の変化形の出現頻度
  - b. 不定形と定形 (3 単現以外) で同じ形式を用いること (ゼロ形態素の使用)

- (54) 一致のみに焦点を当てるため、簡易的な構造で学習を行う
- a. 訓練文  
 (S (NN teacher) (VP (VBZ (VB praise)) (NN student)) (. .))  
 (S (NNS teacher) (VP (VBP (VB praise)) (NNS students)) (. .))
  - b. テスト文  
 (S (NN student) (VP (VBZ (VB praise)) (NN teacher)) (. .))  
 (S (NNS students) (VP (VBZ (VB praise)) (NN teacher)) (. .))
- 主語名詞句の数の違いにより差がみられるか

(55)



- (56) 3単現の PPL が高い  
 → 3単現の一致は機械学習においても誤りが出現することから、言語にその要因がある可能性を示唆する

#### 4. まとめ

**RQ** 3単元の-s脱落の要因は、中間言語の欠陥ではなく、英語の構造に起因するか

- (57) まとめ
- a. 3単元の-s脱落の要因は、中間言語の欠陥ではなく、英語の構造に起因する可能性を示した
  - b. 今後は心理言語学の面でも同様の結果が得られるような実験方式を開発したい

## REFERENCES

- Bird, Steven, Ewan Klein, and Edward Loper (2009) *Natural language processing with Python: analyzing text with the natural language toolkit*, O'Reilly Media, Inc.
- Bobaljik, Jonathan David (2017) "Distributed Morphology," *Oxford Research Encyclopedia of Linguistics* 2017, 1–42.
- Carden, Guy (1983) "The Non-Finite = State-Ness of the Word Formation Component," *Linguistic Inquiry* 14, 537–541.
- Creutz, Mathias, and Crista Lagus. 2007. Unsupervised models for morpheme segmentation and morphology learning. *ACM Transactions on Speech and Language Processing* 4.
- Dyer, Chris, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A. Smith (2016) "Recurrent neural network grammars.," In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 199–209, Association for Computational Linguistics, California.
- Frimu, Rodica and Laurent Dekydtpotter (2022) "Inference-based Form Selection in Native and Non-native Speakers: Evidence from Subject-verb Agreement Error Detection in French," *Lingua* 270, 1–23.
- Halle, Morris and Alec Marantz (1993) "Distributed Morphology and the Pieces of Inflection," in Hale, Ken and Samuel Keyser eds. *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, 111–176, Cambridge, MA.: MIT Press.
- Harley, Heidi and Rolf Noyer (1999) "Distributed Morphology," *Glott International* 4, 3–9.
- Haznedar, Belma and Bonnie D. Schwartz (1997) "Are There Optional Infinitives in Child L2 Acquisition," *Proceedings of The 21st Annual Boston University Conference on Language Development* 21, 257–268.
- Hawkins, Roger, and Sarah Liszka (2001) "Locating the source of defective past tense marking in advanced L2 English speakers," *Language Acquisition and Language Disorders* 30, 21-44.
- Kuncoro, Adhiguna, Miguel Ballesteros, Lingpeng Kong, Chris Dyer, Graham Neubig, and Noah A. Smith (2017) "What do recurrent neural network grammars learn about syntax?," In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics 1*, Long Papers, 1249–1258, Association for Computational Linguistics, Valencia, Spain.
- Kim, Yoon, Alexander Rush, Lei Yu, Adhiguna Kuncoro, Chris Dyer, and Gábor Melis (2019) "Unsupervised Recurrent Neural Network Grammars," In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 1*, Long and Short Papers, 1105-1117, Association for Computational Linguistics, Minneapolis, Minnesota.
- Kirov, Christo and Ryan Cotterell (2018) "Recurrent Neural Networks in Linguistic Theory: Revisiting Pinker and Prince (1988) and the Past Tense Debate," *Transactions of the Association for Computational Linguistics*, 651–665.
- Miller, George A (1995) "WordNet: a lexical database for English," *Communications of the ACM* 38 (11), 39-41.
- Noji, Hiroshi and Yohei Oseki (2021) Effective Batching for Recurrent Neural Network Grammars. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 4340–4352, Association for Computational Linguistics, Online.

- Oseki, Yohei (2018) "Syntactic Structures in Morphological Processing," Ph.D. dissertation, New York University.
- Spencer, Andrew (1991) *Morphological Theory: An introduction to Word Structure in Generative Grammar*, Oxford: Blackwell.
- Prévost, Philippe and Lydia White (2000) "Missing Surface Inflection Hypothesis or Impairment in Second Language Acquisition? Evidence from Tense and Agreement," *Second Language Research* 16(2), 103–133.
- R Core Team (2022) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, Retrieved from <http://www.R-project.org/>.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin (2017) "Attention is all you need," In *Advances in Neural Information Processing Systems* (30), 5998–6008, Curran Associates, Inc.