

文法知識と意味知識の習得過程に着目した 言語モデルの第二言語習得の認知モデルとしての評価

乙川凜 宮本友樹 内海彰

電気通信大学

o2430026@edu.cc.uec.ac.jp {miyamoto,utsumi}@uec.ac.jp

概要

本研究は、日本語を第一言語 (L1) として学習したニューラル言語モデル (LM) が、英語の第二言語 (L2) 習得において人間と同様の発達曲線を示すかどうかを検証した。検証では、文法知識と意味知識において、L1 が L2 習得に及ぼす「負の転移」の有無に焦点を当てた。文法知識では、人間の L2 習得過程に類似した発達曲線が確認されたが、意味知識では、L1 学習に起因する負の転移は認められなかった。この結果から、LM は文法習得に関しては人間の認知モデルとして妥当であるが、意味知識の習得においては妥当とは言えないことが示された。

1 はじめに

近年のニューラル言語モデルの発展により、言語モデル (LM) を人間の認知モデルとして捉える試みが多くなされている。特に、第一言語 (L1) 習得としてのモデルのみならず [1, 2, 3], 第二言語 (L2) 習得のモデルとしての研究も行われている [4, 5, 6]。

LM を認知モデルとして妥当であるかを検証するためには、人間の言語習得に見られる主要な現象が LM においても再現されるかどうか重要な論点となる。人間の L2 習得では、L1 の影響を受けることで学習が促進・抑制される「言語間転移」という現象が存在する。これまで、多くの実証研究によってその現象が確認されている [7, 8]。

Yadavalli ら [4] や大羽ら [5] は、L1 と L2 を学習させた LM が言語間転移を示すかを研究した。しかし、L1 学習後の LM が L2 を習得する過程が、人間の L2 学習者と同様の習得過程を示すかどうかについては、注目されてこなかった。また、先行研究で用いられた文法性判断ベンチマーク (BLiMP) は、文法知識を網羅的に測ることはできるが、L1 の影響によって生じる、人間らしい誤りの傾向 (負の転

移) を捉えるには十分ではない。

そこで本研究では、言語能力を「文法知識」と「意味知識」に分類し、負の言語間転移に着目して人間と LM の L2 習得過程を比較することで、LM の L2 習得の認知モデルとしての妥当性を検証する。

2 言語能力の分類

本研究では、言語能力を「文法知識」と「意味知識」に分類する。これは認知神経科学における Ullman [9] の Declarative/Procedural Model に基づく。Ullman は、人間の言語処理が規則的な文法処理を支える「手続き記憶」と語彙知識を支える「宣言的記憶」の二つの記憶システムで支えられると主張した。本研究では前者に対応する知識を「文法知識」、後者に対応する知識を「意味知識」と呼ぶ。

3 研究目的

本研究の目的は、L2 習得の認知モデルとしての LM の妥当性を検証することである。そのために、前述した「文法知識」と「意味知識」の二つの側面に関して、LM が人間の L2 学習者の習得過程を模倣するかどうかを評価する。

4 第二言語モデルの認知モデル

4.1 本研究で用いる言語モデル

評価対象に XLM [10] という多言語対応モデルを用いる。これは Transformer [11] のエンコーダアーキテクチャを基盤とする。対訳ペアを用いて Translation Language Modeling で学習を行うことで、異なる言語を共通の意味空間で学習することができる。

4.2 認知モデルの構築

L2 として英語を学ぶ日本語母語話者を想定した JA-EN モデルと、英語母語話者を想定した EN-Only

モデルを構築する。JA-EN モデル構築のため、初期化した XLM に対して、CC-100 コーパスの約 2 億語の日本語データを用いて事前学習を行う。その後、tatoeba コーパスから抽出した約 200 万語の英語データと、その日本語訳で追加学習を行う。

EN-Only モデルの構築では、初期化した XLM に対して、前述の英語部分のみを用いて学習を行う。

5 妥当性の検証方法

5.1 文法知識

文法知識の習得において、言語間転移の影響は L2 習得の段階に応じて変化することが考えられる。例えば、日本語には冠詞という文法範疇が存在しないため、英語初学者は "I am student." のように冠詞を欠落させる誤りを犯しやすい。このような初期の単純な誤りは上達と共に減少するが、一方で新たに学んだ表現を使おうとし、新しい間違いが生じる。

本研究では、日本人英語学習者の習熟度別の誤り傾向を反映したデータセットを用いて、LM の文法知識習得における言語間転移の影響を評価する。

5.1.1 データセットの構築

L2 学習者の習熟度別の誤り傾向を反映したデータセット構築のため、日本人英語学習者による英語発話コーパスである NICT JLE コーパス [12] から、学習者が生成した誤文と、その修正文である正文から構成される文ペアを収集した。

このコーパスでは学習者は 9 つの習熟度に分類されているが、本研究ではこれらを 3 つのカテゴリ（Group 1-3：初級レベル、Group 4-6：中級レベル、Group 7-9：上級レベル）にグループ化した。

5.1.2 検証方法

モデルの学習は、学習データ全体を一巡する単位であるエポックごとに行われる。各エポックにおいて、LM に 5.1.1 節の文ペアのどちらが正しいかを判別させる。判別は、LM で Perplexity (PPL) を計算することで行う。PPL は、与えられた文に含まれる各単語の出現確率に基づき、その文の予測の困難さを示す指標である。正しく自然な文は、LM が高い確率で単語を予測できるため、PPL が低い値をとる傾向がある。この性質を利用し、PPL が低い文をモデルが正文と判断したとみなし、それが正しい判断である割合を正解率として算出する。

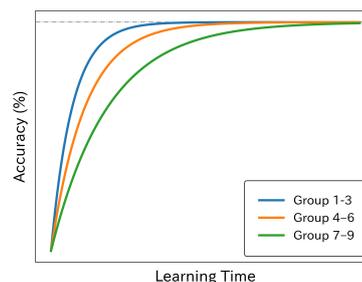


図 1 人間の L2 文法知識習得における理想的な発達曲線

習熟度別の正解率の推移を観察し、モデルが人間の L2 習得過程を模倣するかを評価する。図 1 は、人間の L2 習得における理想的な発達曲線であり、JA-EN モデルが同様の発達曲線を示す場合、L2 習得の認知モデルとして妥当であると判断する。また、EN-Only モデルとの比較により、JA-EN モデルの学習曲線が L1 の影響を受けていることを確認する。

5.2 意味知識

5.2.1 同訳効果

意味知識の習得に関して、L1 において同一の訳語を持つ L2 語彙（同訳ペア）の習得が困難であるという「同訳効果」という現象が知られている [13, 14]。同訳効果の例として、以下のような二つの文を考える。

- I still have strong feelings for John. (正)
- I still have strong emotions for John. (誤)

この例では、前者の "feelings" を用いることが自然であり、後者の "emotions" は不自然である。しかし、日本語では両者とも「感情」と訳されるため、日本語を母語とする英語学習者は適切な語の選択ができず、このような文ペアに対する正解率は低下する。

Jiang [13, 14] は、この現象のメカニズムとして、成人の L2 語彙学習においては L2 単語が L1 のレマ（意味・統語情報）に直接マッピングされる「L1 レマ仲介」が生じていると論じた。さらに Jiang は、この L1 知識への依存は極めて強固であり、学習が進んだ上級者であっても L2 独自の意味体系への再構築は難しく、誤用が持続しやすいことを指摘した。

5.2.2 利用するデータセット

本研究では、同訳効果を検証するための同訳ペアとして、クレシラ [15] が構築したデータセットを用いる。このデータセットは、同一の文脈において、単語選択のみが異なり、一方は自然で、他方は不自

然となる同訳ペアから構築されている。データセットの構築手順は以下の通りである。

1. 候補抽出: 日本語 WordNet より、同一の日本語レンマを持つ英語レンマのペア（同訳ペア）を抽出する。
2. ペア作成: 品詞の一致および語幹の不一致を条件にペアリングを行う。
3. 文作成: 英語 WordNet の用例を正例とし、ペアの相方で置換したものを負例としてペアを作成する。
4. 文法選別: Python ライブラリの LanguageTool を用い、負例に文法的な誤りが生じないものを採用する。
5. 意味選別: 3 種類の BERT 系モデルによる多数決を行い、文ペアを選定する。

最終的に、537 個の同訳ペアからなる 563 問が構築されている。

5.2.3 検証方法

同訳効果が生じる場合、JA-EN モデルの正解率は、L1 知識を持たない EN-Only モデルよりも、学習過程において一貫して低い値を示すと考えられる。

これを検証するため、文法知識の検証の際と同様に、PPL を用いてモデルによる同訳ペアの正誤判定を行う。具体的には、各問題について正文（正しい語彙を含む文）と誤文（同訳ペアの相方で置換した文）の PPL を比較し、正文の PPL の方が低い場合を正解として正解率を算出する。そして、JA-EN モデルの正解率が、EN-Only モデルと比較して一貫して低いかどうかを検証する。

6 検証結果

6.1 文法知識

5.1.1 節のデータセットに対する各エポックでの XLM の二つのモデルの正解率を図 2, 3 に示す。

JA-EN モデルでは、まず初級レベル (Group 1-3) の正解率が学習初期から急速に向上し、70 エポック付近で 75% 程度の正解率に達した後、向上が見られなくなった。次に中級レベル (Group 4-6) は、Group 1-3 よりも緩やかに正解率を伸ばし、100 エポック付近で同様に 75% 程の正解率に収束した。最後に上級レベル (Group 7-9) は、最も緩やかに正解率が向上し、150 エポック付近で他のグループとほぼ同等

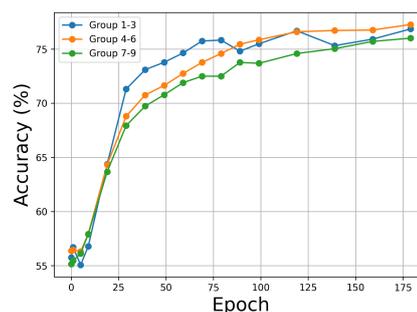


図 2 JA-EN モデルの文法知識の学習曲線

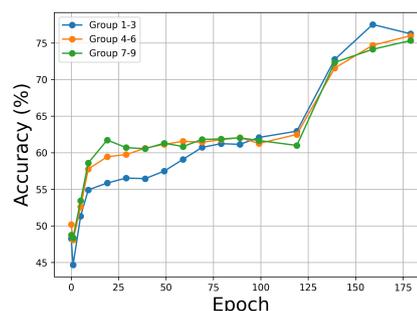


図 3 EN-Only モデルの文法知識の学習曲線

の 75% 程度の正解率に収束した。このように、各グループの正解率の向上にタイムラグが見られ、習熟度に応じて段階的に誤りを克服していることがわかる。これは図 1 に示した人間の L2 習得における理想的な発達曲線の傾向と一致している。

これに対し EN-Only モデルでは、学習開始直後は初級レベルの正解率が低かったものの、学習が進行すると各グループの正解率は同程度になり、学習後半では習熟度別の差異は見られなかった。最終的な正解率は各グループともに約 75% 程度に収束した。

6.2 意味知識

同訳効果の問題に対する、各エポックでの XLM の二つのモデルの正解率を図 4 に示す。学習全体を通じて、JA-EN モデルと EN-Only モデルの正解率に大きな差は見られず、JA-EN モデルが一貫して劣るという傾向は確認されなかった。

7 考察

7.1 文法知識

文法知識で用いた問題は、人間の L2 学習者の習熟度別の誤り傾向を反映している。JA-EN モデルは習熟度に応じて誤りを克服する学習の軌跡が見られたが、EN-Only モデルには見られなかった。これ

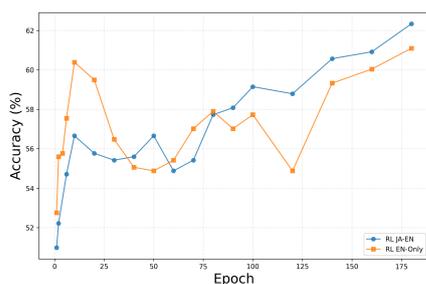


図4 同訳効果の学習曲線

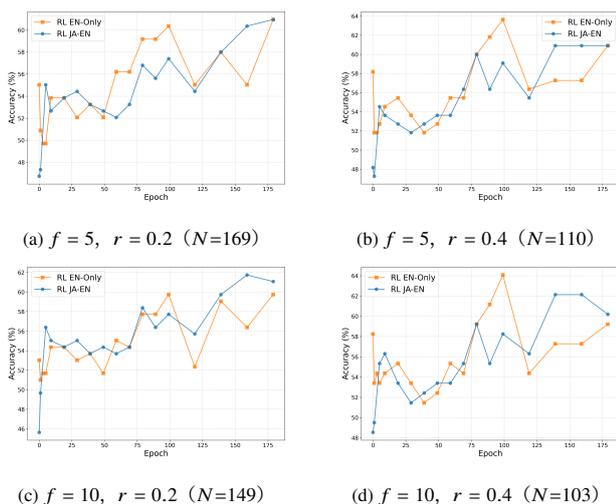


図5 f, r に基づいてフィルタリングした条件による同訳効果の学習曲線

は、JA-EN モデルが L1 の影響を受け、人間の L2 習得における誤りパターンを内在化したことを示す。

習熟度グループ間の正解率に有意差が存在するかを検証するため、フリードマン検定を行い、有意差が認められた場合にはネメニー法による対比較を実施した。JA-EN モデルにおけるフリードマン検定では、習熟度グループ間に有意な差が示された ($p < .001$)。対比較において、正解率は Group7-9 よりも Group1-3 および 4-6 の方が有意に高かった (ともに $p < .001$)。Group 1-3 と 4-6 の間には有意な差は見られなかった。一方で、EN-Only モデルでは、習熟度グループ間に有意な差は見られなかった。

以上の結果から、文法知識において、JA-EN モデルは人間の L2 学習過程を模倣し、認知モデルとしての妥当性を示していると考えられる。

7.2 意味知識

意味知識においては、同訳効果は観察されなかった。この傾向が学習コーパスの特徴に依存するかどうかの追加検証を行った。まず、英語コーパス内の単語の出現回数を f 、同訳ペアの頻度比を r と定義した。ここで r は、出現頻度の低い単語の頻度を

高い単語の頻度で割った値である。低頻度語や極端な頻度差によるノイズを排除するため、同訳ペアの両語がどちらも f 回以上出現し、かつ頻度比が r 以上である条件で同訳ペアのフィルタリングを行い、再評価を実施した。

$f \in \{5, 10\}$ と $r \in \{0.2, 0.4\}$ を組み合わせた 4 条件で再評価を行ったところ図 5 のような結果が得られたが、いずれの条件においても同訳効果は観察されなかった。以上より、意味知識に関して、LM は人間の L2 習得の認知モデルとして妥当とは言えない。

7.3 文法知識と意味知識における対照性

文法知識と意味知識では対照的な結果であった。文法知識では「語順」や「隣の単語との関係」のような構造のルールが支配的である。JA-EN モデルは最初の学習で、日本語特有の構造的なバイアスを強く獲得している。このバイアスの修正は、一度獲得された統語的制約を再構築する必要があるため、容易ではない。ゆえに、英語学習時にも強い干渉を引き起こし、人間同様の誤り傾向が生じたと考えられる。

一方、意味知識では、単語の意味は文脈に応じたベクトルの数値的な微調整によって処理される。人間は L1 概念に強く固着して抜け出せないが、LM において、意味のズレは文脈情報を用いた最適化によって容易に上書き可能であり、結果として人間特有の同訳効果は観察されなかったと考えられる。

8 おわりに

本研究では、文法知識と意味知識の側面から、LM の L2 習得の認知モデルとしての妥当性を検証した。

検証の結果、文法知識の習得では、L1 の知識を持つ JA-EN モデルが人間の学習者と同様に、習熟度に応じて段階的に誤りを克服する過程を示した。これは、LM が人間の L2 文法習得過程を模倣する認知モデルとして、妥当性があることを示している。

一方、意味知識の習得では「同訳効果」は確認されず、LM は人間の L2 意味習得の認知モデルとして妥当とは言えない結果となった。

今後の課題として視覚言語モデル用いた検証が挙げられる。本研究で LM が意味知識において人間と同様の干渉を示さなかった要因として、記号接地の欠如が考えられる。画像などの身体性を伴う情報を統合して学習したモデルでは、概念形成のプロセスが変化し、同訳効果が生じる可能性がある。

参考文献

- [1] Tal Linzen and Marco Baroni. Syntactic structure from deep learning. **Annual Review of Linguistics**, Vol. 7, pp. 195–212, 2021.
- [2] Ariel Goldstein, Zaid Zada, Eliav Buchnik, Mariano Schain, Amy Price, Bobbi Aubrey, Samuel A Nastase, Amir Feder, D Emanuel, A Cohen, et al. Shared computational principles for language processing in humans and deep language models. **Nature Neuroscience**, Vol. 25, pp. 369–380, 2022.
- [3] Alex Warstadt, Leshem Choshen, Aaron Mueller, Adina Williams, Ethan Cheng, Chengxu Zhuang, Jennifer Hu, et al. Findings of the babylm challenge: Sample-efficient pretraining on developmentally plausible corpora. In **Proceedings of the 27th Conference on Computational Natural Language Learning (CoNLL)**, pp. 1–34, 2023.
- [4] Aditya Yadavalli, Alekhya Yadavalli, and Vera Tobin. SLABERT talk pretty one day: Modeling second language acquisition with BERT. In **Proc. of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 11763–11777, 2023.
- [5] 大羽未悠, 栗林樹生, 大内啓樹, 渡辺太郎. 言語モデルの第二言語獲得. 自然言語処理, Vol. 31, No. 2, pp. 433–455, 2024.
- [6] Ionut Constantinescu, Tiago Pimentel, Ryan Cotterell, and Alex Warstadt. Investigating critical period effects in language acquisition through neural language models. **Transactions of the Association for Computational Linguistics**, Vol. 13, pp. 96–120, 2025.
- [7] Job Schepens, Roeland van Hout, and T. Florian Jaeger. Big data suggest strong constraints of linguistic similarity on adult language learning. **Cognition**, Vol. 194, p. 104056, 2020.
- [8] Leida C. Tolentino and Natasha Tokowicz. Cross-language similarity modulates effectiveness of second language grammar instruction. **Language Learning**, Vol. 64, No. 2, pp. 279–309, 2014.
- [9] Michael T Ullman. Contributions of memory circuits to language: The declarative/procedural model. **Cognition**, Vol. 92, No. 1-2, pp. 231–270, 2004.
- [10] G. Lample and A. Conneau. Cross-lingual language model pretraining. In **Advances in Neural Information Processing Systems**, Vol. 32, pp. 7057–7067, 2019.
- [11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. **Advances in neural information processing systems**, Vol. 30, , 2017.
- [12] 情報通信研究機構. The NICT JLE corpus. <https://alaginrc.nict.go.jp/resources/jle/index.html>, 2012. Ver. 4.1.
- [13] Nan Jiang. Form–meaning mapping in vocabulary acquisition in a second language. **Studies in Second Language Acquisition**, Vol. 24, No. 4, pp. 617–637, 2002.
- [14] Nan Jiang. Semantic transfer and development in adult L2 vocabulary acquisition. In Paul Bogaards and Batia Laufer, editors, **Vocabulary in a Second Language: Selection, Acquisition, and Testing**, pp. 101–126. John Benjamins, Amsterdam, 2004.
- [15] クレシアシル, 内海彰, 宮本友樹. 言語モデルを用いた第二言語獲得の認知モデル：意味知識の言語間転移による妥当性の検証. 修士論文, 電気通信大学大学院, 2025. 修士論文.