




どのような言語モデルが不可能な言語を学習してしまうのか？ —語順普遍を例に—

栗林樹生  上田亮  吉田遼  大関洋平 

Ted Briscoe  Timothy Baldwin  

 MBZUAI  東京大学  メルボルン大学

{tatsuki.kuribayashi, ted.briscoe, timothy.baldwin}@mbzuai.ac.ae

{ueda-ryo796, yoshiryo0617, oseki}@g.ecc.u-tokyo.ac.jp

概要

ニューラル言語モデルは、自然言語の範囲を超え、プログラムなど系列データ一般を処理するための基盤的な道具立てとなった。一方言語学では、自然言語のみを受理（学習）し、可能言語と不可能言語を区別する計算モデル（形式言語など）が探求されており、両分野の向かう方向は異なる。そこで本研究では、両視点から、自然言語らしくないデータをモデリングできない言語モデルを探求することで、自然言語の性質に迫る。今回は特に語順に焦点を当て、どのような言語モデルが、自然言語における語順普遍から逸脱した不可能語順をもつデータを学習してしまうのか調査する。認知的に動機づけられた言語モデルが相対的に可能語順のみを学習しやすい傾向にあることなどがわかった。

1 はじめに

ニューラル言語モデルは、系列データを処理するための汎用的な道具立てとなり、様々な言語、プログラム、画像チャンク、さらには自然言語らしくない性質をもつデータ [1] までも学習できてしまう。一方言語学では、自然言語をたらしめる特徴に関心があり、自然言語のみを受理（学習）できる必要十分な計算モデル（形式文法など）が探求されてきた。このような言語学的立場から見れば、いかなる系列データも学習できてしまうニューラル言語モデルは、自然言語そのもののモデルではないと批判される [2]。しかし実際のところ、ニューラル言語モデルがどの程度、自然言語らしくない不可能言語を学習できてしまうのか、また何か工夫を施すことで可能言語のみをうまく学習する言語モデルが構築できるのかといった知見は乏しい。

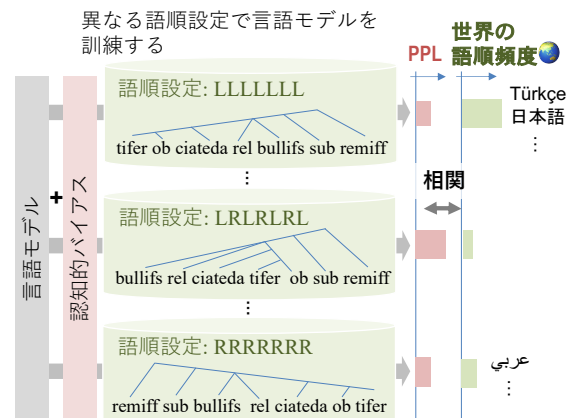


図1 言語モデリングのしやすさによって、自然言語に存在する可能語順と存在しない不可能語順を（柔らかく）区別できるのか調査する。具体的には、典型的に高頻度な自然言語らしい語順ほど言語モデリングがしやすいことを期待する。様々な言語モデルを分析し、どの言語モデルがこの観点で望まれる性質をもつかが焦点となる。

本研究では、自然言語の語順普遍に着目し、どのような言語モデルが語順普遍から逸脱したデータの学習の難しさを創発するのか探求し、構成論的な方法で語順普遍の根源に迫る。具体的には、自然言語に存在する可能語順をもつ人工言語コーパスや、そうでないコーパスをそれぞれ作成し、可能語順のみ低いパープレキシティを示す言語モデルを探す。認知的に動機づけられている言語モデルも含め、20種類の言語モデルの学習（帰納）バイアスについて体系的に調査する。

実験結果から、おおよそ全ての言語モデルにおいて、可能語順は学習しやすい傾向にあることがわかった。これは次の単語を予測しやすいという特性が自然言語の語順を形付けている可能性を示唆し、言語モデリングという枠組みが自然言語に特有な一側面を説明できることを支持している。また通常の言語モデルと認知的に動機づけられた言語モデルで

表 1 語順パラメータと語順の例

パラメータ	L	R
s^S	猫が 食べる。	食べる 猫が。
s^{VP}	猫が 鼠を 食べる。	猫が 食べる 鼠を。
s^{PP}	猫が 机の上で 食べる。	猫がの上で 机 食べる。
s^{NP}	小さな 猫が 食べる。	猫 小さな が食べる。
s^{Rel}	乳を飲む 猫が 食べる。	猫 乳を飲む が食べる。
s^{Case}	猫 が 食べる。	が 猫 食べる。

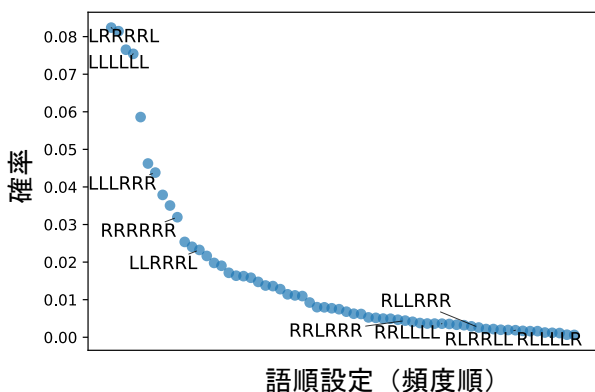


図 2 自然言語における語順設定の頻度分布

は、自然言語の語順分布について異なる側面を創発することがわかった。例えば通常の Transformer 言語モデルは、句横断的に主要部が左か右いずれかに偏るとい調和的な語順を選好することが観察された。一方で認知的に動機づけられた言語モデルの語順選好は、語順の頻度分布全体とより相関しやすい傾向にあった。後者の知見については、自然言語の語順の設計に人間の認知的なバイアスが反映されていること概念実証と見ることができ、計算心理言語学と言語類型論の繋がりを示唆している。

2 実験設定

2.1 自然言語における語順の偏り

本研究では、まず語順に関する 6 つのパラメータ (表 1) を導入し、世界の言語を分類する。各パラメータは特定の種類の句における語順 (枝分かれ方向) を示しており¹⁾、例えば s^S は主語が述語に先行するのかを決定する。以降これらのパラメータの組み合わせを L/R からなる 6 文字の系列で表し、 $[s_i^S, s_i^{VP}, s_i^{PP}, s_i^{NP}, s_i^{Rel}, s_i^{Case}]$ の順に語順設定を表す。例えば LLLLLL は、VP と Case のみ右枝分かれで、他の句では左枝分かれとなる語順設定である。日本語の語順は完全な左枝分かれ型であり LLLLLL と表される。図 2 に、自然言語における $64 = 2^6$ 通りの

1) 主要部に対して従属部が右/左に現れる場合を R/L とする。

語順の頻度分布を示す。The World Atlas of Language Structures (WALS) [3] 中 1,616 言語の情報に基づいて推定したものである。なお 6 つの語順パラメータ全てについて情報が付与されている言語は限られていたため、語順の頻度について単純な平滑化をしている (付録)。特定の語順、特に LLLLLL や RRRRRR といった L と R どちらかが支配的な調和的語順の頻度が高いことがわかる [4]。このような偏りは人間の認知バイアスに紐づくという見方もある [5]。

2.2 言語モデルの語順選好

次に言語モデルの語順選好を定量化する。既存研究で作成された人工言語コーパスの集合 [6] を用いる。このデータは 64 の人工言語コーパス l_k からなり、いずれも人工的に設計した確率文脈自由文法 (PCFG) で作成されており、互いに語順設定のみが異なる。64 コーパスのそれぞれは、§2.1 で定義したそれぞれの語順設定 (LRRRRR など) に対応している。それぞれのコーパスで言語モデルを訓練し、どの設定でパープレキシティ (PPL) が低くなるのかを観察する。この PPL 分布を $p = [PPL(LLLLLL), PPL(LLLLLL), \dots, PPL(RRRRRR)]$ とし、自然言語における語順分布と比較する。なお、全てのコーパスは対応する文法のもとで同じ生成確率をもつため、PPL の違いはモデルの帰納バイアスによってのみ生じる。

2.3 評価尺度

複数の指標を用いて、PPL 分布が可能語順と不可能語順を (柔らかく) 区別できているのか多角的に評価する。

全体相関 PPL 分布 p と、各語順の頻度分布 $f = [\text{freq}(LLLLLL), \text{freq}(LLLLLL), \dots, \text{freq}(RRRRRR)]$ の近さを報告する。具体的には、両者の負のピアソン相関 $-\text{Corr}(p, f)$ を報告する。このスコアが高いほど語順選好の言語類型論的な妥当性が高い。以降 Corr はピアソン相関を表す。

部分相関 部分的な相関として、SOV (LL...), SVO (LR...), OVS (RL...), VOS (RR...) それぞれの部分集合における相関の平均も報告する：

$$-\frac{1}{4}(\text{Corr}(p_{LL}, f_{LL}) + \text{Corr}(p_{LR}, f_{LR}) \quad (1)$$

$$+ \text{Corr}(p_{RL}, f_{RL}) + \text{Corr}(p_{RR}, f_{RR})) \quad (2)$$

ここで p_{XY} および f_{XY} は、 $s^S = X$ かつ $s^{VP} = Y$ を満たす言語に限った PPL および頻度分布である。

可能語順峻別度 不可能な語順設定に対して、可能な語順設定をどれだけ選好できるのかが本研究の重要な論点である。今回用いた 64 語順設定の内、WALS において語順設定の完全一致する言語が存在するものを可能語順 \mathcal{L}_p とし²⁾、それらに対して相対的に良い PPL が割り振られているのかを報告する。具体的には、可能語順の検索問題とみなし、PPL の低い順に 64 語順を並べ、可能語順の平均逆順位を求めた：

$$\frac{1}{|\mathcal{L}_p|} \sum_{l_k \in \mathcal{L}_p} \frac{1}{\text{rank}(l_k)} . \quad (3)$$

調和性 自然言語では L または R のどちらかが支配的な調和的語順が好まれることを踏まえ [4]、言語モデルがそのような語順をモデリングしやすいのかを報告する：

$$- \text{Corr}(p, c) , \quad (4)$$

$$c := [\max(\#L(l), \#R(l)) \text{ for } l \text{ in } L] . \quad (5)$$

ここで、 $\#L(\cdot)$ または $\#R(\cdot)$ は各語順設定における L または R の数であり、両値の最大値は枝分け方向がどれだけ揃っているのかを示す。例えば $\#L(\text{LLLLLL})$ は 6 となる。自然言語における語順分布では、この値は 0.37 である。

左枝分かれ選好度 モデルがどちらかの枝分かれ方向 (L または R) を好む傾向にあるのかも合わせて報告する。具体的には以下の値を報告し、値が高いほど左枝分かれ構造が好まれることを表す：

$$- \text{Corr}(p, b) , \quad (6)$$

$$b := [\#L(l) \text{ for } l \text{ in } L] . \quad (7)$$

自然言語における語順分布では 0.11 となり、どちらかの枝分かれ方向に対する強い選好は存在しない。

2.4 モデル

20 種類の言語モデルを調査する。いずれも単方向言語モデルであり、BPE をコーパスに適用し、サブワード単位で学習されている。具体的には、まず通常の言語モデルとして、Transformer [7]、LSTM [8]、単純再帰型ニューラルネットワーク (SRN) [9]、N-グラム言語モデルを対象とする。代表的な大規模言語モデルとして、LLaMA2 (7B) [10] の few-shot learning による PPL も報告する。さらに、(1) 逐次的統語

2) WALS における報告バイアスを懸念し、対応する個別言語が 1 つなどの可能性境界付近の語順を除き、対応する個別言語が 3 言語以上存在する 8 語順を可能言語とした。

的処理 [11, 12]、(2) 逐次統語解析戦略 [13, 14]、(3) 短期記憶の制限 [15, 16] という 3 つの観点で認知的に動機づけられた言語モデルも分析対象に加える。(1) と (2) の観点から、トップダウン型 (TD) と左隅型 (LC)³⁾ の再帰的ニューラルネットワーク文法 (RNNGs) を用いる [19]。⁴⁾ その理論的な性質により [13]、左隅型のモデルが妥当なバイアスを持つことを期待する。(3) の観点では RNNG の (双方向) LSTM を単純再帰型ニューラルネットワークに置き換えた SRNNG、さらにボトムアップ型 (BU) を含む異なる解析戦略 (TD, BU, LC) の元で作成した逐次統語解析操作系列上で N-グラム言語モデル (統語 N-グラム) を訓練し、短期記憶の制限された統語言語モデルとして対象に加える。なお、統語言語モデルの学習に必要な統語構造情報は、PCFG から文をサンプルする際に同時に手に入る。

3 実験設定

異なるシードを用いて 5 回実験を行い、各指標の平均値及び標準偏差を報告する。各試行では、20,000 文からなるコーパスを作成し、学習・開発・評価セットを 8:1:1 に分割して実験する。可読性のため値を 100 倍して報告する。結果を表 2 に示す。

4 実験結果・考察

言語モデリングのしやすさは語順可能性を創発する まず、全体相関、部分相関、峻別度、調和性において、おおむね全ての言語モデルがチャンスレートよりも自然言語の分布に近い値を示している。このことは自然言語の語順が次単語予測をしやすいように設計されていることを示唆しており、言語の効率性の観点と接続できる [21, 22]。

異なる言語モデルは語順可能性の異なる側面を創発する 一方、特定の言語モデルがすべての指標で特筆した選好を示しているわけでないこともわかる。全体として、Transformer などの通常の言語モデルは峻別度・調和性の観点で高いスコアを示し、一方認知的に動機づけられた言語モデルは語順分布との相関が全体・部分的に高いという傾向が見られる。Transformer が最も自然言語に近い調和性スコアを示していることは、工学的な動機で設計されたモデルが言語類型論的にも尤もらしい選好を創発して

3) 既存研究 [17, 18, 19] に従い、arc-standard アルゴリズムを使用しており、ボトムアップ型との類似性から、左枝分かれ構造を過剰に好むバイアスが反映されている可能性がある。

4) word-synchronous beam search [20] の下で PPL を計算する。

表2 各言語モデルの語順選好. 全体相関, 部分相関, 峻別度における上位3スコアと, 自然言語に最も近い3スコアを太字にしている. また各モデルで最もPPLが低くなった3語順設定も併記する.

モデル	統語	文脈制限	全体相関↑	部分相関↑	峻別度↑	調和性↗	左選好→	上位3言語
自然言語			100.0	100.0	19.1	36.8	10.5	LRRRRL, LRRRRR, LLLRRL
Transformer			12.1 ± 4.3	16.7 ± 6.4	13.0 ± 4.1	37.2 ± 2.3	23.8 ± 6.1	LLLLLL, LLRLLL, RLRLLL
LSTM		✓	10.7 ± 14.7	26.9 ± 7.4	9.5 ± 1.8	23.8 ± 17.8	-1.2 ± 11.3	RLRLLL, RLLLLL, RRRRRR
SRN		✓	16.3 ± 9.4	38.3 ± 3.0	8.2 ± 1.2	16.9 ± 16.3	-3.6 ± 10.4	RLLLLL, RLRLLL, RLRLLL
5-グラム		✓	5.4 ± 1.0	17.0 ± 1.7	13.0 ± 3.0	31.3 ± 0.9	-5.8 ± 1.6	RRLRRR, RRRRRR, LRLRRR
4-グラム		✓	6.5 ± 1.0	16.5 ± 1.6	14.7 ± 4.2	32.7 ± 1.0	-6.4 ± 1.6	RRLRRR, RRRRRR, LRLRRR
3-グラム		✓	8.8 ± 0.7	17.5 ± 1.0	16.9 ± 3.6	34.3 ± 0.7	-6.1 ± 1.0	RRRRRR, RRLRRR, LRLRRR
RNNG	TD		9.4 ± 3.5	-31.5 ± 11.6	4.4 ± 0.8	-7.6 ± 5.1	-30.1 ± 5.7	RRRRLL, RRRLLL, RRLLLL
SRNNG	TD	✓	14.6 ± 8.7	-2.8 ± 5.9	4.5 ± 0.8	-2.0 ± 4.7	-21.5 ± 8.9	LLLLRR, LLRRRR, RLRRRR
RNNG	LC		-23.4 ± 7.0	26.5 ± 13.9	8.9 ± 4.3	-12.7 ± 7.4	-26.2 ± 7.2	RLRLRL, RLRLLL, RRRRRR
SRNNG	LC	✓	17.2 ± 8.5	18.3 ± 12.4	8.9 ± 2.4	5.3 ± 7.9	-36.7 ± 9.7	LRRRRR, LRLRRR, RLRLRR
統語 5-グラム	TD	✓	11.8 ± 2.4	50.4 ± 2.8	6.8 ± 0.5	27.7 ± 1.5	10.2 ± 7.8	RLLRRL, RLRRRL, RLLLLL
統語 4-グラム	TD	✓	29.2 ± 0.6	40.0 ± 1.6	7.9 ± 0.9	28.6 ± 1.7	4.4 ± 5.4	LLRRRL, RLRRRL, LLRRRR
統語 3-グラム	TD	✓	19.9 ± 0.7	29.0 ± 1.8	8.4 ± 2.3	19.4 ± 0.4	17.3 ± 2.0	RLLRRL, LLLRRL, RLLRRR
統語 5-グラム	BU	✓	12.7 ± 1.0	47.9 ± 1.2	4.8 ± 0.0	15.3 ± 0.7	16.3 ± 2.7	RLLRRL, RLRRRL, LLLRLL
統語 4-グラム	BU	✓	20.9 ± 1.1	56.4 ± 0.7	5.1 ± 0.3	13.7 ± 0.6	11.4 ± 2.2	RLLRRL, RLRRRL, LLLRLL
統語 3-グラム	BU	✓	23.1 ± 0.6	52.8 ± 0.2	5.0 ± 0.1	9.6 ± 0.4	-20.1 ± 0.6	RLLRRL, RLLRLL, RLRRRL
統語 5-グラム	LC	✓	18.6 ± 0.9	47.0 ± 1.3	5.5 ± 0.1	20.5 ± 0.6	28.8 ± 1.4	LLLLRLL, RLLRLL, RLRLLL
統語 4-グラム	LC	✓	21.7 ± 0.6	50.6 ± 0.6	5.3 ± 0.1	20.5 ± 0.4	22.2 ± 0.9	RLLRLL, RLRLLL, LLLRLL
統語 3-グラム	LC	✓	18.0 ± 0.2	55.7 ± 0.3	4.7 ± 0.2	13.7 ± 0.3	27.0 ± 0.5	RLLRRR, RLLRRL, RLRRRR
LLaMA2 (7B)			6.9 ± 31.0	15.4 ± 2.5	5.7 ± 2.3	8.2 ± 3.9	-4.6 ± 31.0	LRLLLL, LRLLLL, LRLRLL
チャンスレート			0.0	0.0	7.5	0.0	0.0	-

いる点で興味深い. 一方全体・部分相関では, 統語 N-グラムが自然言語に近い語順選好を示しており, 通常の N-グラム言語モデルよりも相関が向上している点から, **短期記憶の制約と共に統語構造を明示的に考慮する言語モデルの優位性を示唆している.** 言い換えれば, 可能語順は限られた短期記憶で逐次統語解析しやすいと解釈できる. 短期記憶制限の優位性は, RNNG に対する SRNNG の優位性, SRN の全体・部分相関が Transformer や LSTM よりも良いことにも現れている (ただし峻別度などの観点では一部結果が逆転している). なお, 統語解析戦略の違いに関して, 特に統語 N-グラムモデルにおいては明確な差が観察されなかった. また, RNNG に関しても解析戦略に関わらず右枝分かれ言語を好むなど, 必ずしも用いた戦略の理論的な性質 [13] がモデルの振る舞いに現れないことも確認された.

Transformer による調和的語順の創発 Transformer の調和的語順への選好と, 文脈内学習の成功との接続可能性について議論する. 既存研究 [23] では, Transformer が, 同じ関数 (関係) が異なる要素に繰り返し適用される文法 (例えば, **日本 東京 中国 北京 UAE アブダビ** のような国-首都関係) の推定に長けており, それが文脈内学習成功の鍵であるとしている. Transformer がこのような文法を推定できるのであれば, LLLLLL や RRRRRR といった, 修飾

語-被修飾語あるいは被修飾語-修飾語といった同じ関係のペアが一貫した順序で繰り返し出現する傾向にある調和的な語順 (**食べる 急いで ご飯 たくさん 人 隣** のなど) の単純さに気づき, うまく言語モデリングできるのかもしれない.

{S, O, V} 語順分布の創発の難しさ 全体相関の値が部分相関に大きく劣ることから, 言語モデルは主語・目的語・動詞の並び (s^S と s^{VP} の設定) の自然言語らしさを創発できていない. 今回用いた人工言語では, 主語と目的語は同様の名詞句集合からサンプルされており, 一方自然言語では情報の新旧や句の長さなど両者に様々な非対称性が存在する [24, 25, 26]. この限界の原因がモデルにあるのか人工言語データにあるのかは, 人工言語をより自然言語に近づけていくことで切り分ける必要がある.

5 おわりに

本研究では, 言語学的な動機から, 自然言語らしくないデータを学習できない言語モデルについて探求した. 特に自然言語における語順設定の偏りに焦点を当て, 自然言語で頻発する語順設定の言語をモデリングしやすい言語モデルの探索を通して, 語順の偏りが生じる要因について構成論的な検証を試みた. より自然言語に近いデータを用い, 自然言語の理解へ知見を還元していくことを展望としたい.

謝辞

本研究は、JST さきがけ JPMJPR21C2 の支援を受けたものです。

参考文献

- [1] Jeff Mitchell and Jeffrey Bowers. Priorless recurrent networks learn curiously. In Donia Scott, Nuria Bel, and Chengqing Zong, editors, **Proceedings of the 28th International Conference on Computational Linguistics**, pp. 5147–5158, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics.
- [2] N Chomsky, I Roberts, and J Watumull. Noam chomsky: The false promise of ChatGPT. **NY Times**, 2023.
- [3] Matthew S. Dryer and Martin Haspelmath, editors. **WALS Online (v2020.3)**. Zenodo, 2013.
- [4] Joseph H Greenberg, et al. Some universals of grammar with particular reference to the order of meaningful elements. **Universals of language**, Vol. 2, pp. 73–113, 1963.
- [5] Jennifer Culbertson, Paul Smolensky, and Géraldine Legendre. Learning biases predict a word order universal. **Cognition**, Vol. 122, No. 3, pp. 306–329, March 2012.
- [6] Jennifer C. White and Ryan Cotterell. Examining the inductive bias of neural language models with artificial languages. In **Proceedings of ACL-IJCNLP 2021**, pp. 454–463, August 2021.
- [7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, L ukasz Kaiser, and Illia Polosukhin. Attention is All you Need. **NIPS**, pp. 5998–6008, 2017.
- [8] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. **Journal of Neural Computation**, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [9] Jeffrey L Elman. Finding structure in time. **Cogn. Sci.**, Vol. 14, No. 2, pp. 179–211, March 1990.
- [10] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajwal Bhargava, Shrusti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Es-iobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovitch, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. Llama 2: Open foundation and Fine-Tuned chat models. **arXiv preprint**, July 2023.
- [11] Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A Smith. Recurrent neural network grammars. In **Proceedings of NAACL 2016**, pp. 199–209, June 2016.
- [12] John Hale, Chris Dyer, Adhiguna Kuncoro, and Jonathan R. Brennan. Finding Syntax in Human Encephalography with Beam Search. In **Proceedings of ACL 2018**, pp. 2727–2736, 2018.
- [13] Philip Resnik. Left-Corner parsing and psychological plausibility. In **Proceedings of COLING 1992**, 1992.
- [14] Hiroshi Noji and Yusuke Miyao. Left-corner transitions on dependency parsing. In **Proceedings of COLING 2014**, pp. 2140–2150, 2014.
- [15] Richard Futrell, Edward Gibson, and Roger P. Levy. Lossy-Context Surprisal: An Information-Theoretic Model of Memory Effects in Sentence Processing. **Journal of Cognitive Science**, 2020.
- [16] Tatsuki Kuribayashi, Yohei Oseki, Ana Brassard, and Kentaro Inui. Context limitations make neural language models more human-like. In **Proceedings of EMNLP 2022**, pp. 10421–10436, December 2022.
- [17] Adhiguna Kuncoro, Chris Dyer, John Hale, Dani Yogatama, Stephen Clark, and Phil Blunsom. LSTMs can learn Syntax-Sensitive dependencies well, but modeling structure makes them better. In **Proceedings of ACL 2018**, pp. 1426–1436, July 2018.
- [18] Ryo Yoshida, Hiroshi Noji, and Yohei Oseki. Modeling Human Sentence Processing with Left-Corner Recurrent Neural Network Grammars. In **Proceedings of EMNLP 2021**, pp. 2964–2973, November 2021.
- [19] Hiroshi Noji and Yohei Oseki. Effective batching for recurrent neural network grammars. In **Findings of ACL-IJCNLP 2021**, pp. 4340–4352, August 2021.
- [20] Mitchell Stern, Daniel Fried, and Dan Klein. Effective inference for generative neural parsing. In **Proceedings of EMNLP 2017**, pp. 1695–1700, September 2017.
- [21] Edward Gibson, Richard Futrell, Steven P Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen, and Roger Levy. How efficiency shapes human language. **Trends Cogn. Sci.**, Vol. 23, No. 5, pp. 389–407, May 2019.
- [22] Michael Hahn, Dan Jurafsky, and Richard Futrell. Universals of word order reflect optimization of grammars for efficient communication. **Proc. Natl. Acad. Sci. U. S. A.**, Vol. 117, No. 5, pp. 2347–2353, February 2020.
- [23] Michael Hahn and Navin Goyal. A theory of emergent In-Context learning as implicit structure induction. **arXiv e-prints**, March 2023.
- [24] John A Hawkins. **A performance theory of order and constituency**. No. 73. Cambridge University Press, 1994.
- [25] Jeanette K Gundel. Universals of topic-comment structure. **Studies in syntactic typology**, Vol. 17, No. 1, pp. 209–239, 1988.
- [26] Elisabeth Verhoeven. Thematic asymmetries do matter! a corpus study of german word order. **Journal of Germanic Linguistics**, Vol. 27, No. 1, pp. 45–104, 2015.
- [27] Matthew S. Dryer. Order of subject and verb (v2020.3). In Matthew S. Dryer and Martin Haspelmath, editors, **The World Atlas of Language Structures Online**. Zenodo, 2013.
- [28] Matthew S. Dryer. Order of object and verb (v2020.3). In Matthew S. Dryer and Martin Haspelmath, editors, **The World Atlas of Language Structures Online**. Zenodo, 2013.
- [29] Matthew S. Dryer. Order of adposition and noun phrase (v2020.3). In Matthew S. Dryer and Martin Haspelmath, editors, **The World Atlas of Language Structures Online**. Zenodo, 2013.
- [30] Matthew S. Dryer. Order of adjective and noun (v2020.3). In Matthew S. Dryer and Martin Haspelmath, editors, **The World Atlas of Language Structures Online**. Zenodo, 2013.
- [31] Matthew S. Dryer. Order of relative clause and noun (v2020.3). In Matthew S. Dryer and Martin Haspelmath, editors, **The World Atlas of Language Structures Online**. Zenodo, 2013.
- [32] Matthew S. Dryer. Position of case affixes (v2020.3). In Matthew S. Dryer and Martin Haspelmath, editors, **The World Atlas of Language Structures Online**. Zenodo, 2013.

表3 MALS データの統計

All languages in WALS	2,679
Targeted languages	1,616
Targeted switches	9,696 (=1,616×6)
Missing switches	3,343
LL ratio	46.7%
LR ratio	34.3%
RL ratio	3.6%
RR ratio	15.5%
s^S	82A Order of Subject and Verb [27]
s^{VP}	83A Order of Object and Verb [28]
s^{PP}	85A Order of Adposition and Noun Phrase [29]
s^{NP}	87A Order of Adjective and Noun [30]
s^{Rel}	90A Order of Relative Clause and Noun [31]
s^{Case}	51A Position of Case Affixes [32]

A 語順分布

表3に語順分布の推定に用いたWALSデータの詳細を示す。収録されている2,679言語のうち、少なくとも1つの関連する語順パラメータが記録されている1,616言語の統計を用いた。特定の語順パラメータの情報が欠落している場合は、記録されている他の語順情報と互換性のある語順設定候補全てに対して、1票を分割して投票した。例えば、LXLXR(Xが欠損値)の場合、LLLLL, LLLLRR, LRLLL, and LRLRRに1/4票投票した。

B 人工言語

表4に、人工言語コーパス生成時に用いた文法を記す。平均文長は11.8単語であり、統語構造の平均的な深さは9.1であった。語彙には、既存研究で作成された擬似単語を使っており、合計で1,314種類の単語が用意されている。

表4 人工言語コーパス生成時に用いた文法

確率 導出規則	関連パラメータ
1 ROOT → S	
1/2 S → NP.Subj.S VP.S	s^S
1/2 S → NP.Subj.P VP.P	s^S
1/3 VP.S → VP.Past.S	
1/3 VP.S → VP.Pres.S	
1/3 VP.S → VP.Comp.S	
1/3 VP.P → VP.Past.P	
1/3 VP.P → VP.Pres.P	
1/3 VP.P → VP.Comp.P	
1/2 VP.Comp.S → VP.Comp.Pres.S	
1/2 VP.Comp.S → VP.Comp.Past.S	
1/2 VP.Comp.P → VP.Comp.Pres.P	
1/2 VP.Comp.P → VP.Comp.Past.P	
1/2 VP.Past.S → IVerb.Past.S	
1/2 VP.Past.S → NP.Obj TVerb.Past.S	s^{VP}
1/2 VP.Pres.S → IVerb.Pres.S	
1/2 VP.Pres.S → NP.Obj TVerb.Pres.S	s^{VP}
1/2 VP.Past.P → IVerb.Past.P	
1/2 VP.Past.P → NP.Obj TVerb.Past.P	s^{VP}
1/2 VP.Pres.P → IVerb.Pres.P	
1/2 VP.Pres.P → NP.Obj TVerb.Pres.P	s^{VP}
1 VP.Comp.Pres.S → S.Comp Verb.Comp.Pres.S	s^{VP}
1 VP.Comp.Past.S → S.Comp Verb.Comp.Past.S	s^{VP}
1 VP.Comp.Pres.P → S.Comp Verb.Comp.Pres.P	s^{VP}
1 VP.Comp.Past.P → S.Comp Verb.Comp.Past.P	s^{VP}
1 S.Comp → S Comp	s^{Comp}
1 NP.Subj.S → NP.S Subj	s^{Case}
1 NP.Subj.P → NP.P Subj	s^{Case}
1/2 NP.Obj → NP.S Obj	s^{Case}
1/2 NP.Obj → NP.P Obj	s^{Case}
5/21 NP.S → Noun.S	
5/21 NP.S → Adj Noun.S	s^{NP}
5/21 NP.S → VP.S Rel Noun.S	s^{Rel}
5/21 NP.S → Pronoun.S	
1/21 NP.S → PP NP.S	s^{PP}
10/43 NP.P → Noun.P	
10/43 NP.P → Adj Noun.P	s^{NP}
10/43 NP.P → VP.P Rel Noun.P	s^{Rel}
10/43 NP.P → Pronoun.P	
2/43 NP.P → PP NP.P	s^{PP}
1/172 NP.P → NP.S CC NP.S	
1/172 NP.P → NP.P CC NP.P	
1/172 NP.P → NP.P CC NP.S	
1/172 NP.P → NP.S CC NP.P	
1/2 PP → NP.S Prep	s^{PP}
1/2 PP → NP.P Prep	s^{PP}
1/43 Adj → Adj CC Adj	
1/566 TVerb.Past.S → TVerb.Past.S CC TVerb.Past.S	
1/566 TVerb.Pres.S → TVerb.Pres.S CC TVerb.Pres.S	
1/566 IVerb.Past.S → IVerb.Past.S CC IVerb.Past.S	
1/566 IVerb.Pres.S → IVerb.Pres.S CC IVerb.Pres.S	
1/566 TVerb.Past.P → TVerb.Past.P CC TVerb.Past.P	
1/566 TVerb.Pres.P → TVerb.Pres.P CC TVerb.Pres.P	
1/566 IVerb.Past.P → IVerb.Past.P CC IVerb.Past.P	
1/566 IVerb.Pres.P → IVerb.Pres.P CC IVerb.Pres.P	
1 Verb.Comp.Past.S → word ~ Dict[Verb.Comp.Past.S] # 22 types	
1 Verb.Comp.Past.P → word ~ Dict[Verb.Comp.Past.P] # 22 types	
565/566 IVerb.Past.S → word ~ Dict[IVerb.Past.S] # 113 types	
565/566 IVerb.Past.P → word ~ Dict[IVerb.Past.P] # 113 types	
565/566 TVerb.Past.S → word ~ Dict[TVerb.Past.S] # 113 types	
565/566 TVerb.Past.P → word ~ Dict[TVerb.Past.P] # 113 types	
1 Verb.Comp.Pres.S → word ~ Dict[Verb.Comp.Pres.S] # 22 types	
1 Verb.Comp.Pres.P → word ~ Dict[Verb.Comp.Pres.P] # 22 types	
565/566 IVerb.Pres.S → word ~ Dict[IVerb.Pres.S] # 113 types	
565/566 IVerb.Pres.P → word ~ Dict[IVerb.Pres.P] # 113 types	
565/566 TVerb.Pres.S → word ~ Dict[TVerb.Pres.S] # 113 types	
565/566 TVerb.Pres.P → word ~ Dict[TVerb.Pres.P] # 113 types	
1 Noun.S → word ~ Dict[Noun.S] # 162 types	
1 Noun.P → word ~ Dict[Noun.P] # 162 types	
1 Pronoun.S → word ~ Dict[Pronoun.S] # 5 types	
1 Pronoun.P → word ~ Dict[Pronoun.P] # 2 types	
42/43 Adj → word ~ Dict[Adj] # 42 types	
1 Prep → word ~ Dict[Prep] # 4 types	
1 CC → da	
1 Comp → sa	
1 Rel → rel	
1 Subj → sub	
1 Obj → ob	