

# 意味と頻度の関係から探求する言語モデルの特性

永田亮<sup>1</sup> 今泉允聡<sup>2</sup> 高村大也<sup>3</sup> 川崎義史<sup>2</sup> 大谷直輝<sup>4</sup>

<sup>1</sup> 甲南大学 <sup>2</sup> 東京大学 <sup>3</sup> 産業技術総合研究所 <sup>4</sup> 東京外国語大学  
nagata-nlp2026 @ ml.hyogo-u.ac.jp.

## 概要

本稿では、単語頻度と語義の豊富さに関する経験則 meaning-frequency law を通じて、従来とは異なる切り口で言語モデルの特性を分析できることを報告する。具体的には、(1) 類似した言語を学習済みであれば、言語モデルは未学習言語の語義をある程度識別できる可能性があること、(2) 大規模な言語モデルでは、同法則からは乖離してヒンジ現象という特異な現象が生じることを示す。

## 1 はじめに

本稿では、meaning-frequency law [1] と呼ばれる言語の統計的法則を通じて、従来とは異なる切り口で言語モデルの特性を分析できることを示す。Meaning-frequency law とは、高頻度な単語ほど語義数が多いという経験則である。形式的には、単語頻度を  $f$ 、単語の語義の豊富さを  $v$  としたとき、

$$\log(v) = \delta \log(f) + c \quad (1)$$

が成り立つという法則である。文献 [2] は、言語モデルから得られる文脈付き単語ベクトル（以降、単に単語ベクトルと略記）のばらつきに基づいて語義の豊富さ  $v$  を定量化する手法を提案している。同文献は、そのようにして求めた  $v$  から得た傾き  $\delta$  が言語モデルの語義識別能力に強く相関することを示している。また、言語モデルが未学習の言語では式 (1) からの乖離が大きいことも示している。本稿は、この知見を言語モデルの分析に利用する。

分析の準備として、4 節で、傾き  $\delta$  が、ある条件下で、理論値 0.5 に収束することを証明する。この値は、以降の節で、各種条件下での  $\delta$  の値を吟味する際の一つの基準となる。

続く 5 節では、言語モデルが未学習の言語を対象にしても同法則が観測されることを示す。未学習の言語とは、言語モデルの訓練データに含まれない言語のことである。例えば、クロアチア語やセルボ・

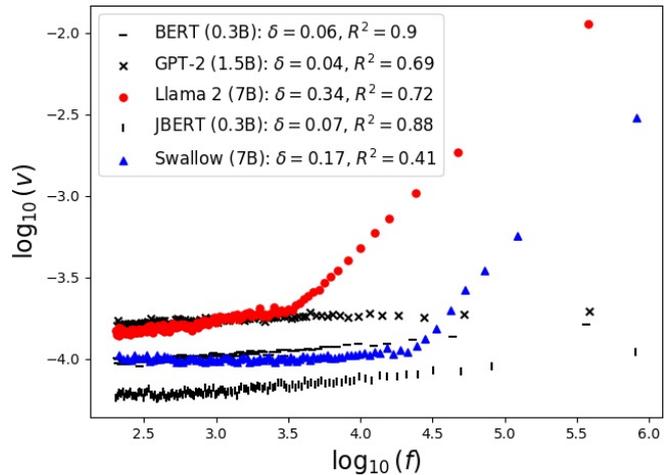


図1 頻度と語義の豊富さの関係におけるヒンジ現象。

クロアチア語を学習した言語モデルを未学習言語ではあるが類縁関係にあるスロヴェニア語に適用して得た語義の豊富さ  $v$  も式 (1) に従うことを示す。この結果は、人間と同様に、言語モデルでも習得済み言語に類似した未学習言語の語義をある程度識別できる可能性があることを示唆する。

6 節では、ヒンジ現象と呼ばれる現象を報告する。ヒンジ現象とは、言語モデルが大規模になると、頻度と語義の豊富さのプロットが単一の直線ではなく、図1の赤丸と青三角で表されるように、二つの直線をつないだヒンジ形となる現象を指す。この現象は、大規模な言語モデルでは高頻度語とそれ以外の語で語義の識別方法が異なることを示唆する。本稿では、英日のデータにおいてヒンジ現象が生じることを示し、その理由を考察する。

## 2 関連研究

本研究は、言語学で知られる meaning-frequency law に依拠する。同法則は、英語 [1, 3]、トルコ語 [4]、日本語 [5] など様々な言語で成り立つことが知られている。これらの従来研究では、語義の豊富さとして辞書に掲載された語義数を用いる。

Nagata らの検証 [2] では、言語モデルから得られる単語ベクトルの方向のばらつきに基づいて語義

の豊富さ  $v$  を定量化する。単語ベクトルの方向は、周辺単語すなわち文脈により決定される。したがって、単語ベクトルの方向がばらつくということは、対応する単語が多様な文脈で使用されることを意味し、更には、語義が豊富であると解釈される。

ベクトルの方向のばらつき（厳密には、その逆の概念である集中度）は、von Mises-Fisher 分布 [6] を通じて定量化が可能である。この分布の確率密度関数は、 $d$  次元の単位ベクトル  $\mathbf{x}$  に対して、 $f(\mathbf{x}; \boldsymbol{\mu}, \kappa) \propto \exp(\kappa \boldsymbol{\mu}^T \mathbf{x})$  と定義される。ここで、 $\boldsymbol{\mu}$  ( $\|\boldsymbol{\mu}\| = 1$ ) と  $\kappa$  ( $\kappa \geq 0$ ) は、それぞれ平均方向と集中度と呼ばれるパラメータである。この分布では、 $\mathbf{x}$  は平均方向を中心に集中度  $\kappa$  で等方的に分布すると考える。言い換えれば、 $\kappa$  は、ベクトルの方向の集中度を表している。本稿では、文献 [2] に従い集中度の逆数  $v \equiv 1/\kappa$  を語義の豊富さと定義する。集中度  $\kappa$  の最尤推定は、近似的に、

$$\kappa \approx \frac{l(d-l^2)}{1-l^2}, \quad (2)$$

になることが知られている [6]。ここで、 $l$  は、単語タイプごとに単語ベクトルを平均した結果のノルムである。定数  $d$  はベクトルの次元である。

文献 [2] は、傾き  $\delta$  の大きさが、語義識別能力と高い正の相関をみせることを報告している。また、モデルサイズが小さい場合や未学習の言語では式 (1) からの乖離が大きくなることも示している。本稿では、この知見を言語モデルの分析に利用する。

### 3 Meaning-frequency law の検証方法

検証は、文献 [2] の方法を踏襲する。具体的には、単語頻度  $f$  と語義の豊富さ  $v = 1/\kappa$  を式 (1) にあてはめ回帰を行う。ただし、回帰の際に、説明変数と目的変数の bin 平均化を行う。すなわち、頻度の大きい順に  $f$  および  $v$  をいくつかの bin にまとめ平均値を求めてから回帰を行う。本稿では、文献 [2] にならい bin サイズ 100 を用いる。従来研究では、回帰の当てはまり具合を表す決定係数  $R^2$  の大きさにより同法則の成り立ち度合いを考える。加えて、傾き  $\delta$  に対する  $p$  値も判断材料としている。本稿でもこの二つの基準を採択する。

具体的な検証方法は次の 7 ステップからなる：(1) 入力コーパス中の単語頻度  $f$  をカウント；(2) 言語モデルを用いて各単語の事例を単語ベクトルに変換；(3) 得られた単語ベクトルのノルムが 1 となるように正規化；(4) 単語タイプごとに平均ベクトル

とそのノルムを算出；(5) 式 (2) を用いて語義の豊富さ  $v = 1/\kappa$  を算出；(6) bin 化した  $f$  と  $v$  に対して両対数で回帰を行い、決定係数  $R^2$ 、傾き  $\delta$  を算出；(7) 決定係数  $R^2$ 、傾き  $\delta$  を散布図とともに出力し、meaning-frequency law が成り立つかを吟味。

## 4 ランダムベクトルにおける傾き

本節では、式 (1) の傾き  $\delta$  の基準としてある種の理論値を与える。ここでは単語列ではなく、直接、単語ベクトルの分布を考える。3 節で述べたように、本稿では、単語ベクトルは von Mises-Fisher 分布に従い  $d-1$  次元の超球面上に分布する。この超球面上に、付録 A.1 に示す手順でランダムにベクトルを配置した際の傾き  $\delta$  の値を考える。この値は、経験的には、0.5 となることが知られている [2]。

理論値は次のように導出できる。設定として、いまいくつかの単語タイプ  $t$  があり、対応するランダムな単語ベクトルがそれぞれ von Mises-Fisher 分布  $P_t$  に従うとする。 $P_t$  は集中度  $\kappa_t$  と平均方向  $\boldsymbol{\mu}_t$  を持ち、この値は  $t$  によって異なる。また  $P_t$  に従うランダムベクトルの期待値を  $\mathbf{e}_t \in \mathbb{R}^d$  とする。ここで  $\|\mathbf{e}_t\| \in (c, 1-c)$  がある  $c \in (0, 0.5)$  のもとで成立すると仮定する。例えば  $c = 0.2$  の場合は  $\|\mathbf{e}_t\| \in (0.2, 0.8)$  になる。この仮定は  $P_t$  が一定程度の広がりを持つときに正当化される。

次に、 $P_t$  から生成された単語ベクトルを用いて、 $\kappa_t$  およびその逆数  $1/\kappa_t$  (すなわちタイプ  $t$  を持つ単語ベクトルの語義の豊富さ) を推定する状況を考える。いま、コーパスが  $N$  個の単語から成り、 $t$  の相対頻度  $f_t \in (0, 1)$  のもとで  $f_t N$  個の単語ベクトルがタイプ  $t$  を持つとする。この  $f_t N$  個の単語ベクトルの経験平均を  $\hat{\mathbf{e}}_t \in \mathbb{R}^d$  とする。このとき、式 (2) より、タイプ  $t$  の条件付き分布の集中度の推定量は  $\kappa_t = \|\hat{\mathbf{e}}_t\|(d - \|\hat{\mathbf{e}}_t\|^2)/(1 - \|\hat{\mathbf{e}}_t\|^2) \approx C\|\hat{\mathbf{e}}_t\|$  である。最後の近似は、上記の経験平均の区間の仮定のもとで、ある定数  $C > 0$  を用いて成立する。ここで、観測した単語ベクトルが  $P_t$  に従うランダムベクトルであることを用いて、中心極限定理より  $\|\hat{\mathbf{e}}_t\| - \|\mathbf{e}_t\| = O((f_t N)^{-1/2})$  が高確率で成立する。なお中心極限定理は平均の推定誤差がガウス分布に収束することを示すもので、 $P_t$  が von Mises-Fisher 分布であることは影響しない。これを用いると、 $t$  の語義の豊富さは  $v_t = 1/\kappa_t \approx 1/(C\|\hat{\mathbf{e}}_t\|) = 1/\{C(\|\mathbf{e}_t\| + O((f_t N)^{-1/2}))\} = O((f_t N)^{1/2})$  と書ける。両辺で対数を取ることで、

$\log(v_i) \approx \frac{1}{2} \log(f_i N) = \frac{1}{2} \log(f)$  が成立する.

## 5 未学習言語における傾き

本節では、言語モデルが未学習の言語データにおいても meaning-frequency law が観測される場合があることを示す. 詳細は付録 A.2 に示すが、検証の概要は次のとおりである. 言語モデルとして multilingual BERT (mBERT) を用いる. 言語データは聖書コーパスとする. 収録言語のうち mBERT が未学習かつ、最低単語頻度、bin サイズ共に 100 としたときに 8 点以上プロットがとれる 5 言語を対象とする. 比較のため、学習済み言語の日本語、英語、またそれぞれシャッフルしたデータ (付録 B) も対象とする.

表 1 に結果を示す. 表 1 より、エスペラント語、スロヴェニア語、ソマリ語において、傾き  $\delta$  と決定係数の値が高いことがわかる. この結果は、未学習の言語では、傾きも決定係数も値が小さくなるという文献 [2] の知見に反する<sup>1)</sup>. エスペラント語は、ヨーロッパ言語をベースにした人工言語であり、それらのベース言語の多くが mBERT の訓練データに含まれている. 同様に、スロヴェニア語と類縁関係にあるクロアチア語、セルボ・クロアチア語も訓練データに含まれている. これら類似言語の学習により未学習言語においても傾き  $\delta$  と決定係数の値が高くなると我々は考えている. ソマリ語についても、同じ語族であるアフロ-アジア語族の言語データが mBERT の訓練データに含まれているため同様な議論が成り立つ可能性はある. ただし、ソマリ語が属するクシ語派の他の言語は含まれておらず、エスペラント語やスロヴェニア語とはかなり状況が異なる. なお、学習済み言語を含めてすべての言語において傾き  $\delta$  の値は 4 節で与えた理論値 0.5 よりかなり小さいことも特徴的である (次の 6 節でみる傾きの値とは対照的である).

上述の予想を検証するために、未学習言語における mBERT の性能を直接評価することが考えられるが、その方法は自明ではない. mBERT はマスク型言語モデルであるので、穴埋め問題を解かせることで、ある言語の理解度を評価することができる. その場合でも、語彙に登録されていない単語は出力できないため、穴に該当する単語が語彙に含まれるように緻密にテストを設計する必要がある. 自己回帰

1) 例えば、語順をシャッフルした英語コーパスでは、傾きが  $-0.36$ 、決定係数が  $0.66$  と報告されている.

表 1 未学習言語における回帰の結果.

言語	傾き $\delta$ ( $p$ 値)	決定係数 $R^2$
Esperanto	0.032 (0.03)	0.79
Slovene	0.031 (0.01)	0.72
Somali	0.019 (0.02)	0.63
Shona	0.002 (0.01)	0.03
Xhosa	0.000 (0.98)	0.00
Shuffled English	0.028 (0.03)	0.71
Shuffled Japanese	0.001 (0.05)	0.07
English	0.042 (0.03)	0.84
Japanese	0.035 (0.07)	0.91

型の言語モデルであれば、与えられた文脈以降の単語列を推定することでこの問題は理論上は回避できる. しかしながら、未学習言語の単語列を出力するには訓練されておらず、この方法も困難を伴うことが予想される. 一方、meaning-frequency law を切り口とした提案方法は、個別の単語ではなく語義の大域的な性質に基づくため、これらの問題を回避し未学習言語における言語モデルの語義識別能力を吟味する手段を提供するともいえる.

以上の結果の解釈には注意も必要である. 対象としたのは聖書コーパスであり特殊なジャンルの文章である. また、サイズも、文献 [2] で用いられている CCOHA などのコーパスに比べて  $1/100$  程度である. 更に、シャッフルした英語コーパスについても、スロヴェニア語と同程度の傾きと決定係数を示すことにも注意が必要である. サイズがより大きい CCOHA をシャッフルしたコーパスでは、頻度と語義の豊富さのプロットが V 字になることが報告されており [2]、傾向がかなり異なる. より確かな知見とするため、聖書コーパス以外の各国語コーパスを用いた検証が必要である.

## 6 ヒンジ現象

本節では、大規模言語モデルにおける傾き  $\delta$  を調査する. 詳細は、付録 A.3 に示すが、検証の概要は次のとおりである. 言語モデルは、Llama2 (英語) と Swallow (日本語) とする (いずれも 7B サイズ). 比較のために、BERT (0.3B)、GPT-2 (1.5B) も用いる. 対象コーパスは、CCOHA [7] の 2000 年代の文書 (英語) と青空文庫コーパス (日本語) とする.

結果は、1 ページ目に示した図 1 のとおりである. 同図より、7B サイズの言語モデルのプロットは他と傾向が大きく異なることがわかる. すなわ

ち、高頻度帯で傾きが劇的に大きくなり、ヒンジ現象がみられる。このように途中で折れ線傾向がみられることはクロスオーバーとして知られており、例えば、中国語の文字の順位頻度分布でみられる [8]。ただし、順位頻度分布は、単語や文字など表層形と頻度を対象にしたものであり、また、傾きが途中で減衰する。我々の知る限り、語義と頻度の関係において図 1 のようなクロスオーバーがみられることを報告するのは本稿が初めてである。この現象は、式 (1) が予想する直線からは明らかに逸脱しており、大規模言語モデルにおいては、高頻度語とそれ以外の語で語義の識別方法が異なる可能性を示唆する。

興味深いことに、低頻度帯ではモデルサイズにかかわらず同程度の傾きである。言い換えれば、低頻度帯では、モデルサイズにかかわらず単語頻度に応じて単語ベクトルは同程度にばらついている。対照的に、大規模なモデルの高頻度帯では傾きが大きくなる。また、言語の違い（英日）と言語モデルの違い（Llama 2, Swallow）にもかかわらず、7B サイズのモデルでは、高頻度帯も傾きが同程度であるのも興味深い。更に、高頻度帯の傾きは 4 節で求めたランダムベクトルに対する理論値 0.5 を超えている；Llama 2 で頻度 5000 以上、Swallow で頻度 20000 以上の範囲の傾きを求めたところ、それぞれ 0.87 と 0.93 であった。このことは、超球面上にランダムにベクトルが生成される状況よりも、単語頻度に対する語義の豊富さの増加が早いことを意味している。ただし、Swallow は Llama 2 を継続学習しているため、ヒンジ現象や傾きにおける類似性は Llama 2 のアーキテクチャに起因している可能性もある。

ヒンジ現象が生じる一つの理由として言語的要因が考えられる。例えば、機能語と内容語で言語モデル内で処理が異なり、得られる単語ベクトルも異なるということが考えられる。残念ながら、機能語と内容語という単純な分類では図 1 のヒンジをうまく説明できない。使用した英語コーパス中で頻度上位 1000 件の単語を分析すると、頻度 100 あたりで既に people, new, think など通常は内容語に分類される語が現れる。それ以外の範囲も目視で確認したところ、英日とも、傾きの違いを説明する分類基準は見つけられなかった。今のところ意味のある分類基準は見つけられていないが、より詳細な分析によりヒンジ現象をうまく説明するものが見つかるかもしれない（場合によっては、言語的に意味のある新たな分類基準が見つかるかもしれない）。

別の要因として、モデル的な要因も考えられる。我々の予想としては、高頻度語では多様体仮説<sup>2)</sup>が低頻度語より成り立ちにくいいためヒンジ現象が生じると考えている。言語モデルの訓練の際には、高頻度な訓練事例（すなわち高頻度語）における訓練ロスを重点的に減らすことが、訓練全体のロスを減らすことにつながる。そのためには、より多くの次元を利用して高頻度語に対する言語モデルの表現力を高めることが効果的である。十分にモデルサイズが大きければ、より頻度が低い単語についても表現力を高められるが、実際にはモデルサイズは有限であり、表現力は制限される。その結果、より頻度が低い単語については表現力が不足し、対応する単語ベクトルは低次元の多様体に押し込める必要がある。以上が我々が考えている仮説である。

この仮説が正しいとすると、言語モデルのサイズが大きくなるにつれてヒンジの折れ目が低頻度側に移動していくという予想ができる。最終的には、折れ目がなくなり、再び一本の直線になる可能性もある。また、高頻度語とそれ以外の語について、それぞれのベクトル集合の固有次元を求める検証方法も考えられる。仮説が正しいければ、高頻度語のほうが固有次元が多くなるはずである。ただし、単語ベクトルの個数は非常に多く、また次元数も大きい（Llama 2 では 4096 次元）ため計算量的な難しさが存在する。計算量の問題を解決して、固有次元を求めることが今後の研究の一つの方向性となる。

## 7 おわりに

本稿では、meaning-frequency law を通じて、従来とは異なる切り口で言語モデルの特性を分析できることを報告した。まずは、準備としてランダムベクトルを用いた場合、同法則の判断基準となる直線の傾き  $\delta$  が理論値 0.5 に収束することを証明した。次に、同法則を利用して、人間と同様に、言語モデルでも習得済み言語に類似した未学習言語の語義をある程度識別できる可能性があることを示した。更に、7B サイズの大規模な言語モデルでは、meaning-frequency law からは乖離して、単語頻度と語義の豊富さのプロットがヒンジ型になることも報告した。ヒンジ現象は、大規模な言語モデルでは、高頻度語とそれ以外の語で語義の識別方法が異なることを示唆している。

2) 多様体仮説とは、高次元空間上のベクトルで表されるデータが、実際にははるかに次元の低い多様体上に分布しているという仮説である。

## 謝辞

この成果の一部は、産総研政策予算プロジェクト「フィジカル領域の生成 AI 基盤モデルに関する研究開発」および JSPS 科研費 JP23K12152 の助成により得られたものである。産総研及び AIST Solutions が提供する ABCI3.0 を「ABCI 3.0 開発加速利用」支援を受けて利用した。

## 参考文献

- [1] George Kingsley Zipf. The meaning-frequency relationship of words. **The Journal of General Psychology**, Vol. 33, No. 2, pp. 251–256, 1945.
- [2] Ryo Nagata and Kumiko Tanaka-Ishii. A new formulation of Zipf’s meaning-frequency law through contextual diversity. In **Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 15323–15335, 2025.
- [3] Philip Edmonds. **Lexical Disambiguation**, pp. 43–62. Elsevier, Amsterdam, 2005.
- [4] Bahar Ilgen and Bahar Karaoglan. Investigation of Zipf’s ‘law-of-meaning’ on turkish corpora. In **Proceedings of the 22nd International Symposium on Computer and Information Sciences**, pp. 1–6, 2007.
- [5] Francis Bond, Arkadiusz Janz, Marek Maziarz, and Ewa Rudnicka. Testing Zipf’s meaning-frequency law with wordnets as sense inventories. In **Proceedings of the 10th Global WordNet Conference**, pp. 342–352, 2019.
- [6] Arindam Banerjee, Inderjit S. Dhillon, Joydeep Ghosh, and Suvrit Sra. Clustering on the unit hypersphere using von Mises-Fisher distributions. **Journal of Machine Learning Research**, Vol. 6, No. 46, pp. 1345–1382, 2005.
- [7] Reem Alatrash, Dominik Schlechtweg, Jonas Kuhn, and Sabine Schulte im Walde. CCOHA: Clean corpus of historical American English. In **Proc. of the 12th Language Resources and Evaluation Conference**, pp. 6958–6966, 2020.
- [8] 田中久美子. 言語とフラクタル. 東京大学出版会, 東京, 2021.
- [9] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners, 2019.

## A 検証の詳細設定

### A.1 ランダムベクトルの生成方法

4 節でランダムベクトルに対する傾き  $\delta$  の理論値を議論した. その手順は次のとおりである. (1) コーパス中の単語頻度をカウント, (2) 各単語の頻度に応じた個数のランダムベクトルを超球面上に生成, (3) 得られた単語ベクトルから  $v$  を求め, 単語頻度に回帰することで  $\delta$  の値を算出. なお, ステップ (2) では  $d$  次元の標準正規分布に従いサンプリングしたベクトルを正規化することで,  $d-1$  次元の超球面上にランダムに生成できることに注意されたい. このようにして生成したランダムベクトルに対するプロットは図 2 のようになる.

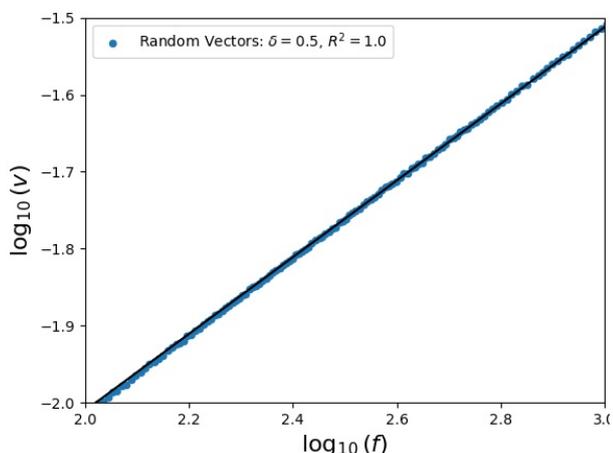


図 2 ランダムベクトルにおける頻度と語義の豊富さの関係.

### A.2 未学習言語に関する検証の詳細

言語モデルとして, multilingual BERT (bert-base-multilingual-cased<sup>3)</sup>) を用いた. また, 多言語パラレルコーパスである聖書コーパス<sup>4)</sup>を対象データとした. 聖書コーパスにおいて multilingual BERT が未学習である言語<sup>5)</sup>のうち, 最低単語頻度を 100, bin サイズを 100 としたときに, 8 点以上プロットがとれる 5 言語を対象とした. 具体的な言語は, 表 1 に示すとおりである. なお, トークン分割は multilingual BERT のトークナイザに従った.

- 3) <https://huggingface.co/google-bert/bert-base-multilingual-cased>
- 4) <https://github.com/christos-c/bible-corpus>
- 5) 学習済みの言語は次のページで公開されている: <https://github.com/google-research/bert/blob/master/multilingual.md#list-of-languages>

### A.3 ヒンジ現象に関する検証の詳細

言語モデルとして, Llama2 (Llama-2-7b-chat-hf<sup>6)</sup>; 7B サイズ; 英語) と Swallow (Swallow-7b-instruct-hf<sup>7)</sup>; 7B サイズ; 日本語) を用いた.

また, 比較のために, BERT (bert-large-uncased<sup>8)</sup>; 0.3B サイズ; 英語), GPT-2 [9] (gpt2-xl<sup>9)</sup>; 1.5B サイズ; 英語), 日本語 BERT (bert-large-japanese-v2<sup>10)</sup>; 0.3B サイズ) も用いた (図 1 の凡例では JBERT と表記した).

対象コーパスは, 英語については CCOHA [7] の 2000 年代の文書, 日本語については青空文庫コーパス<sup>11)</sup>とした. トークン分割は, 各言語モデルのトークナイザを使用した.

## B シャッフルコーパスの詳細

シャッフルコーパスの作成方法は次のとおりである. 入力コーパスを一つの長い単語列とみなし, ランダムに 1 単語ずつ選び出して, 入力コーパスと同一の長さの新たな単語列を作り出した. この手順を 10 回繰り返し, シャッフルコーパスとした. 得られた単語列中の記号.?! (日本語の場合は。)を文末位置とみなした. こうして得られるシャッフルコーパスは, 単語の並びはランダムにシャッフルされるが, 長さおよび各単語の頻度の点で元コーパスと同一となる.

- 6) <https://huggingface.co/meta-llama/Llama-2-7b-chat-hf>
- 7) <https://huggingface.co/tokyotech-llm/Swallow-7b-instruct-hf>
- 8) [https://huggingface.co/docs/transformers/model\\_doc/bert](https://huggingface.co/docs/transformers/model_doc/bert)
- 9) <https://huggingface.co/openai-community/gpt2>
- 10) <https://huggingface.co/tohoku-nltp/bert-large-japanese-v2>
- 11) [https://github.com/aozorahack/aozorabunko\\_text](https://github.com/aozorahack/aozorabunko_text). 2023 年 12 月 3 日アクセス.