

# 日本語の読みやすさに対する 情報量に基づいた統一的な解釈

栗林 樹生<sup>1,2</sup> 大関 洋平<sup>3,4</sup> 伊藤 拓海<sup>1,2</sup> 吉田 遼<sup>3</sup> 浅原 正幸<sup>5</sup> 乾 健太郎<sup>1,4</sup>

<sup>1</sup> 東北大学 <sup>2</sup> Langsmith 株式会社 <sup>3</sup> 東京大学 <sup>4</sup> 理化学研究所 <sup>5</sup> 国立国語研究所  
 {kuribayashi, t-ito, inui}@ecei.tohoku.ac.jp  
 {oseki, yoshiryo0617}@g.ecc.u-tokyo.ac.jp, masayu-a@ninjal.ac.jp

## 1 はじめに

読みやすい文章とはどのようなものだろうか？文章を書いて情報伝達をする全てのヒトに関わる問いである。本研究では、「読む」というヒトの情報処理のモデルについてモデルが何を計算しているかというレベルで理解し、文章の読みやすさについて洞察を得ることを目指す。

### 1.1 背景

英語では、特に言語モデルによって計算されるサプライザル ( $-\log p(\text{単語や文} | \text{先行文脈})$ ) が単語や文の逐次的な処理負荷をうまくモデリングできるというコーパス全体の傾向が報告されてきた [1, 2, 3, 4, 5, 6, 7, 8]。ヒトは続きを予測しながら文章を読んでおり、予想と異なる情報が登場する（サプライザルが高くなる）と処理負荷が高まる（例えば、読み時間が長くなる）というサプライザル理論に基づく [9, 10]。一方日本語では、近年読み時間データ (BCCWJ-EyeTrack) が整備され [11]、日本語の逐次的な読みやすさについて探索的な分析が行われてきた [11, 12, 13, 14]。例えば「係り元が先行文脈に多く存在する文節が読みやすい」 [11]、「聞き手にとって既知と想定される情報は読みやすい」といった記述がされてきた [14]。これらの記述は、ヒトの文処理モデル解明の手がかりとなるだけでなく、読みやすい文章を書く上での知見としても役立つ可能性がある。しかしながら、これらの個別の記述間の共通点については定かではなく、新たな傾向の探索は依然困難である。

### 1.2 仮説と検証結果

サプライザル理論に基づき、BCCWJ-EyeTrack において記述的研究で得られてきた日本語読み時間に

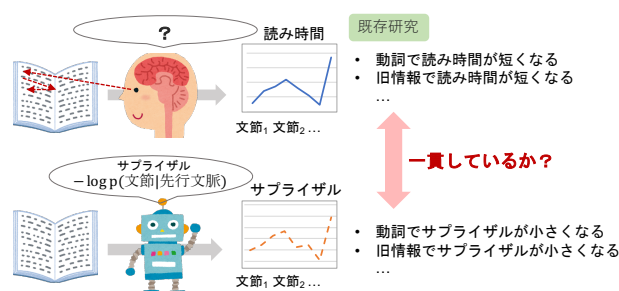


図1 ヒトの読み時間と言語モデルから得られるサプライザルの傾向の比較の概要。

関する傾向が、サプライザルが大きいく所読みにくくなるという傾向として統一的に解釈できるという仮説を立てる。この仮説が支持されれば、読み時間に対する個別の観察間に繋がりが生まれ、例えば観測の難しいヒトに現れる現象（読み時間の変化など）ではなく文章に付与されたサプライザルを分析することで、読みやすさの傾向に関する新たな記述に繋がる可能性がある。このような記述は、サプライザルが大きいく所読みにくくなるといった実際のライティングに適用しにくい知見を噛み砕いたものとして有益であると考えられる。また「なぜある言語現象が読み時間に影響を与えるのか」といったヒトに対する現象個別の科学的な問いが「なぜサプライザルがヒトの読み時間に影響を与えるのか」というヒトに対する一般的な問いと「なぜその言語現象がコーパス上の確率に影響を与えるのか」というデータで検証しやすい問いに整理される。さらに、サプライザル理論については英語コーパス上での検証が中心であり、日本語における本理論の一般性についても示唆を与える。加えて、工学的応用として言語モデルを文章の逐次的な読みやすさの評価に用いる妥当性の検証にも繋がる。

Transformer 自己回帰言語モデルを用いた実験より、少なくとも BCCWJ-EyeTrack 上で報告されてきた統語から談話レベルに及ぶ読み時間の様々な

傾向についてはサプライザル上でも概ね再現された。すなわち、個別の記述が情報量の観点から統一的に解釈できることが支持された。またヒトは先を予想しながら文章を読んでおり、予想できない情報が現れると読み負荷が大きくなるというサプライザル理論の言語横断的な一般性も支持された。最後に読み時間とサプライザルの相違点について分析し、特に PPL の低い言語モデルで、ヒトの読み時間で得られていた特定の観察が極端なバイアスとして観測される傾向を報告する。本研究で BCCWJ-EyeTrack に対して付与したサプライザルデータ (BCCWJ-UniSeqLM) は公開する<sup>1)</sup>。

## 2 実験設定

### 2.1 言語モデル

文章レベルの left-to-right 言語モデルを訓練した。入力サブワードに分割した<sup>2)</sup>。読み時間は文節ごとに付与されており、長さ  $N$  の文  $w_{1:N} = (w_1, w_2, \dots, w_N)$  内の長さ  $T$  の文節  $b = w_{k:k+T}$  ( $1 < k \leq k+T \leq N$ )<sup>3)</sup> について、文節を構成するサブワードの同時確率のサプライザル  $-\log_2 p(w_k, \dots, w_{k+T} | w_1, \dots, w_{k-1}) = -\sum_{i=k}^{k+T} \log_2 p(w_i | w_1, \dots, w_{i-1})$  を計算した。ハイパーパラメータや学習データ量に関しては予備分析を行い、PPL が最も低い設定 (LM<sub>PPL</sub>) と、サプライザルによってヒトの読み時間を最もうまくモデリングできた設定 (LM<sub>PSY</sub>) の 2 種類を実験に用いる<sup>4)</sup>。予備分析については付録に記載する。

### 2.2 サプライザルのモデリング

それぞれの言語モデルを用いて BCCWJ-EyeTrack 上の各文節にサプライザルを付与し、既存研究 [12, 13, 14] でヒトの読み時間に対して行われていた分析を、被説明変数を読み時間からサプライザルに置き換えて実施した。具体的には、既存研究に従い下式の右辺に分析ごとに興味の対象となる説明変数 (例えば統語範疇ラベルなど) を加えてサプライ

ザルをモデリングし、回帰係数の傾向を分析した：

$$\text{surprisal} \sim \text{length} + \text{BOS} + \text{EOS} + \text{pre\_EOS} + \text{sentN} \\ + \text{tokenN} + (1|\text{article}) + (1|\text{seed}) . \quad (1)$$

各言語モデルについて異なるシードで 3 つのモデルを学習し、記事 ID と言語モデルのシードを変量因子として扱った。回帰モデルには一般化線形混合モデルを用いた。

スペースの都合上、実験ではサプライザルをモデリングした結果の回帰係数の符号および、係数が 0 であるという対立仮説に対する p 値 (\*p<0.1; \*\*p<0.05; \*\*\*p<0.01) を報告する。また分析に応じて、係数同士の大小関係も補助的に報告する<sup>5)</sup> (結果の表中で背景色を灰色にして記載する)。またヒトの読み時間についても、式 1 の被説明変数を読み時間に置き換えて同様の検証を行い、サプライザルをモデリングした場合と係数の傾向を比較する。読み時間としては、視線走査法により計測された対数注視時間 (first pass time, FPT) と自己ペース読文法により計測された対数読み時間 (SELF) を用いた [11]<sup>6)</sup>。いずれの分析においても、読み時間が 0 であるか 3 標準偏差を超えるデータポイントはあらかじめ除外し、被説明変数を標準化した。最終的に 13,148 のデータポイントを用いた。

## 3 実験

### 3.1 統語・意味範疇

コーパス中の各文節には、統語範疇ラベル (体, 用, 相の類など) と意味範疇ラベル (関係, 主体, 活動, 生産物, 自然の部門など) が付与されており、(i) 読み時間が体の類>相の類>用の類となる傾向、(ii) 関係の部門の読み時間が短くなる傾向が報告されている [12]。既存研究に従い各文節の統語・意味範疇のラベルを式 1 の右辺に固定因子として追加し、サプライザル・読み時間をモデリングした。

ヒトの FPT, SELF と、LM<sub>PPL</sub>, LM<sub>PSY</sub> から得られたサプライザルについて、回帰係数の一部を表 1 に示す。例えば「用 (v.s. 体)」の行ではヒトと言語モデルいずれにおいても符号が負であることより、体の類よりも用の類に分類される文節の方がサプライ

1) <https://github.com/kuribayashi4/BCCWJ-UniSeqLM>  
 2) mecab [15] と unidic で国語研短単位に分割した後、バイト対符号化を用いてさらに分割した (character coverage=0.9995, vocab size=100000).  
 3) BOS トークン ( $w_1$ ) の存在から、文節は  $w_2$  以降から始まる。  
 4) 学習データは新聞記事と日本語 Wikipedia (1.4G サブワード) から成り、LM<sub>PPL</sub> の学習にはすべてのデータを、LM<sub>PSY</sub> では 1/10 の学習データを使った。また、LM<sub>PPL</sub> に対して LM<sub>PSY</sub> は 1/100 のアップデート回数で訓練され、およそ 1/10 のパラメータ数である。

5) 大小関係の統計的有意性については、固定因子 A, B に対する回帰係数の推定量の差  $|\alpha_A - \alpha_B|$  が、 $\alpha_A$  と  $\alpha_B$  のいずれにおいても 90% (\*), 95% (\*\*), 99% (\*\*\*) 信頼区間を超えるかで判断する。

6) 式 1 の被説明変数を読み時間に置き換え、スクリーン上の表示位置などの説明変数を加えた。詳細は付録に記述する。

表1 読み時間・サプライザルと、意味・統語範疇の関係。係数の差については背景色を灰色にする。

固定因子		FPT	SELF	LM <sub>PSY</sub>	LM <sub>PPL</sub>
統語	用 (vs. 体)	***	***	***	***
	相 (vs. 体)	***	***	***	***
	分類不可 (vs. 体)	-	*	**	***
	用 (vs. 相)	**	-	***	***
意味	主体 (vs. 関係)	***	+	-	-
	活動 (vs. 関係)	***	**	**	-
	生産物 (vs. 関係)	**	***	**	+
	自然 (vs. 関係)	+	**	***	+

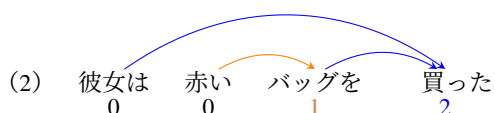
表2 読み時間・サプライザルと anti-locality 効果。

固定因子	FPT	SELF	LM <sub>PSY</sub>	LM <sub>PPL</sub>
掛かり元数	***	***	***	***

ザルが小さく、読み時間も短くなる傾向にあることが分かる。統語範疇においては、読み時間とサプライザルいずれも体の類>相の類>用の類の順に小さくなるという一貫した傾向がみられた。意味範疇においては、関係の部門の読み時間が短くなるという読み時間の傾向が LM<sub>PSY</sub> で概ね再現された。

### 3.2 統語構造

日本語をはじめとする主辞後置言語では、先行文脈に係り元が多く存在するほど読み時間が短くなる anti-locality 効果が報告されている [16, 17]。例えば下の例では、「買った」の係り元は2つ存在し「バッグを」については1つ存在するため、「買った」の方が読み時間が短くなるという効果である。



式1の右辺に、先行文脈に存在する文節の係り元数を固定因子として加え、サプライザルと読み時間をモデリングした。表2より、係り元数に対する係数が有意に負であることから、読み時間とサプライザルいずれにおいても係り元数が多くなるほど処理負荷が低くなるという anti-locality 効果が確認された。本結果はサプライザル理論が anti-locality 効果を説明できるとする予想 [10] を支持する。

### 3.3 節境界

英語では節末で読み時間が長くなる wrap-up 効果が報告されているが [18, 19]、日本語では wrap-up 効果が生じないという報告がされてきた [13]。式1の

表3 読み時間・サプライザルと wrap-up 効果。

固定因子	FPT	SELF	LM <sub>PSY</sub>	LM <sub>PPL</sub>
節末	***	***	***	***

表4 読み時間・サプライザルと節の種類。

固定因子 (節名)	FPT	SELF	LM <sub>PSY</sub>	LM <sub>PPL</sub>
補足語修飾 (vs. 内容)	-	-	***	-
名詞 (vs. 引用)	-	*	**	*

右辺に、文節が節末に存在するかという固定因子を加え、サプライザルと読み時間をモデリングした。表3より、節末に配置される文節はサプライザルが小さいため読みやすく、読み時間においても wrap-up 効果が生じないことが示唆された。

さらに読み時間では、節の種類によって節末の読み時間の傾向が異なることが報告されてきた [13]。例えば、(i) 内容節 (節内の述語と係り先の語とに述語項関係がない) よりも、補足語修飾節 (節内の述語と係り先の語とに述語項関係がある) の方が節末の読み時間が短い、(ii) 引用節よりも名詞節の方が読み時間が短いといった傾向が観察されている。式1の右辺に、文節が属する節の種類を固定因子として加えモデリングした結果を表3に示す。傾向 (i)、(ii) についてサプライザルでも同様の傾向が観察され、節末ごとのサプライザルの大小の傾向の違いに起因する読み時間の差であることが示唆された。

### 3.4 情報構造

情報構造と日本語の読み時間を照らし合わせた分析が行われてきた [20, 14]。既存研究 [14] に従い、情報構造に直接的・間接的に関わる、(i) 書き手の立場で考える情報の新旧 (情報状態)、(ii) 読み手の立場も含めて考える共有性、(iii) 指示対象を聞き手が同定できるかを示す定性、(iv) 対象が生きているかを示す有生性の4つの観点で分析を行う。英語など冠詞が存在する言語と比べて日本語では情報の性質の表出が少なく、文章に表出されない情報構造によるバイアスがサプライザルに現れるかは、言語モデルの分析という観点でも興味深い。

既存研究に従い、式1の右辺に情報状態、共有性、定性、有生性に対応する固定因子を同時に加え、サプライザルと読み時間をモデリングした。表5より、SELFや既存研究で示されていた (i) 新情報より旧情報の方が読みやすい (情報状態における新情報>旧情報)、(ii) 聞き手と共有できている情報のほうが読みやすい (共有性における非共有>想定可能>

表5 読み時間・サプライザルと情報構造の関係。係数の差については背景色を灰色にする。

固定因子	FPT	SELF	LM <sub>PSY</sub>	LM <sub>PPL</sub>
新 (vs. 旧)	-	+**	+***	+***
共有 (vs. 分類不可)	-	-	+	-
非共有 (vs. 分類不可)	+	+	+***	+
想定可能 (vs. 分類不可)	-	+	+***	-
非共有 (vs. 想定可能)	-	-	-	-
想定可能 (vs. 共有)	-	-	-	-
定 (vs. 不定)	-	+***	+***	+***
有生 (vs. 無生)	-*	+***	-	+***

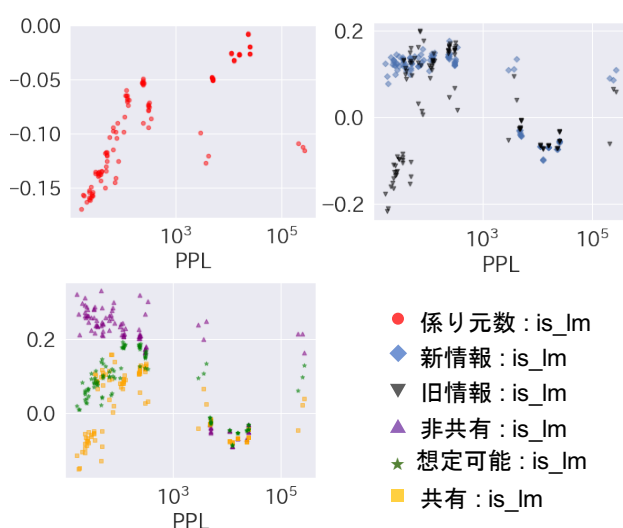


図2 言語モデルのPPLと、その言語モデルから得られたサプライザルを用いてサブグループ解析をした際の各交互作用項に対する回帰係数との関係。

共有), (iii) 定名詞句よりも不定名詞句の方が読みやすい, (iv) 有生名詞句よりも無生名詞句の方が読みやすいといった傾向が、特にLM<sub>PPL</sub>で示された。

#### 4 読み時間とサプライザルの相違点

最後に、ある因子がサプライザルと読み時間に与える影響の差について交互作用の解析を通して調査する。具体的には、サプライザルと読み時間を言語モデルとヒトという異なるグループから得られた「読み負荷」という同一の変数とみなし、読み負荷に対して回帰を行う。本研究では、既存研究で典型的に分析されるFPTを用いた。なお、読み時間とサプライザルはそれぞれあらかじめ平均0、分散1に標準化している。説明変数として「言語モデルから得られた読み負荷であるか」(is\_lm)という二値素性(言語モデルであれば1、ヒトであれば0)を追加し、さらに式1の右辺や分析の対象となる各固定因子それぞれについてis\_lmとの交互作用項を追加す

る。ある因子Aとis\_lmの交互作用項に対する係数( $\alpha_{A:is\_lm}$ )は、因子Aが被説明変数に与える影響が、言語モデルグループとヒトグループでどう異なるかを意味する。例えば $\alpha_{A:is\_lm}$ が有意に0より大きい場合は、ヒトよりも言語モデルに対して因子Aの与える影響が、因子Aの値が大きくなることで読み負荷が増大する方向に大きいことを表す。3節の分析で用いたそれぞれの固定因子について、is\_lmとの交互作用を分析した。また訓練設定が異なる111の言語モデルについてそれぞれ分析を行い、言語モデルのPPLと交互作用の関係を見る(詳細は付録へ)。

特に顕著な傾向が見られた(i) 係り元数(anti-locality効果), (ii) 情報状態, (iii) 共有性に対応する固定因子について、is\_lmとの交互作用項の傾向を報告する(図2)。各プロットが、各言語モデルを用いた実験結果(交互作用項の係数)に対応し、横軸が実験で用いた言語モデルのPPL、縦軸が係数の値である。係り元数、情報状態、共有性いずれにおいてもPPLが低くなるほど係数が0から離れていき、読み時間において観察されていた傾向が過剰に示される方向(anti-localityを支持、新情報>旧情報、非共有>想定可能>共有)にヒトから逸脱していく傾向が見られた。サプライザルで観察されたバイアスからヒトの読み負荷に対して洞察を得る場合、バイアスの大きさについてはサプライザル上で誇張されている可能性があるが、バイアスの方向(ある現象が処理負荷を増加させるか低下させるか)については洞察を得られる可能性が示唆された。

#### 5 おわりに

日本語の読み時間について探索的な分析から記述されてきた傾向について、情報量に基づいて統一的に解釈できるという仮説を立てた。統語から談話レベルの観点において、読み時間とサプライザルの間に様々な共通点が観察され、仮説が支持された。より大規模なデータに基づいた調査<sup>7)</sup>や、脳活動データなどを用いた調査も視野に入りたい。

謝辞. 本研究はJSPS科研費JP20J22697, 19H04990の助成を受けたものです。また、国立国語研究所共同研究プロジェクト「大規模コーパスを利用した言語処理の計算心理言語学的研究」の支援を受けたものです。

7) 特に述語項構造・共参照との関係については、データポイントの少なから今後の検証課題とした。

## 参考文献

- [1]Stefan L. Frank and Rens Bod. Insensitivity of the Human Sentence-Processing System to Hierarchical Structure. *Journal of Psychological Science*, Vol. 22, No. 6, pp. 829–834, 2011.
- [2]Victoria Fossum and Roger Levy. Sequential vs. Hierarchical Syntactic Models of Human Incremental Sentence Processing. In *Proceedings of CMCL*, pp. 61–69, Montréal, Canada, 6 2012. Association for Computational Linguistics.
- [3]Tal Linzen. What can linguistics and deep learning contribute to each other? *Journal of Language*, Vol. 95, No. 1, pp. 99–108, 2019.
- [4]Adam Goodkind and Klinton Bicknell. Predictive power of word surprisal for reading times is a linear function of language model quality. In *Proceedings of CMCL2018*, pp. 10–18, 2018.
- [5]Christoph Aurnhammer and Stefan Frank. Comparing gated and simple recurrent neural network architectures as models of human sentence processing. In *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, pp. 112–118, 2019.
- [6]Ethan Gotlieb Wilcox, Jon Gauthier, Jennifer Hu, Peng Qian, and Roger Levy. On the Predictive Power of Neural Language Models for Human Real-Time Comprehension Behavior. In *Proceedings of CogSci*, pp. 1707–1713, 2020.
- [7]Danny Merx and Stefan L. Frank. Comparing Transformers and RNNs on predicting human sentence processing data. *arXiv preprint arXiv:2005.09471*, 2020, 2020.
- [8]Martin Schrimpf, Idan Blank, Greta Tuckute, Carina Kauf, Eghbal A Hosseini, Nancy Kanwisher, Joshua Tenenbaum, and Evelina Fedorenko. The neural architecture of language: Integrative reverse-engineering converges on a model for predictive processing. *bioRxiv*, 2020.
- [9]John Hale. A Probabilistic Earley Parser as a Psycholinguistic Model. In *Proceedings of NAACL*, pp. 159–166, 2001.
- [10]Roger Levy. Expectation-based syntactic comprehension. *Journal of Cognition*, Vol. 106, No. 3, pp. 1126–1177, 2008.
- [11]Masayuki Asahara, Hajime Ono, and Edson T Miyamoto. Reading-Time Annotations for “Balanced Corpus of Contemporary Written Japanese”. In *Proceedings of COLING*, pp. 684–694, 2016.
- [12]Masayuki Asahara and Sachi Kato. Between Reading Time and Syntactic / Semantic Categories. In *Proceedings of IJCNLP*, pp. 404–412, 2017.
- [13]Masayuki Asahara. Between reading time and clause boundaries in Japanese - wrap-up effect in a head-final language. In *Proceedings of PACLIC*, pp. 19–27, 2018.
- [14]Masayuki Asahara. Between Reading Time and the Information Status of Noun Phrases. *Natural Language Processing*, Vol. 25, No. 5, pp. 527–554, 2018.
- [15]Taku Kudo. Mecab: Yet another part-of-speech and morphological analyzer. <http://mecab.sourceforge.jp>, 2006.
- [16]Lars Konieczny. Locality and parsing complexity. *Journal of Psycholinguistic Research*, Vol. 29, No. 6, pp. 627–645, 2000.
- [17]Shodai Uchida, E Miyamoto, Yuki Hirose, Yuki Kobayashi, and Takane Ito. An ERP study of parsing and memory load in Japanese sentence processing – A comparison between left-corner parsing and the Dependency Locality Theory. *Technical report of IE-ICE. Thought and language*, Vol. 114, pp. 101–106, 2014.
- [18]Marcel A. Just and Patricia A. Carpenter. A theory of reading: From eye fixations to comprehension. *Journal of Psychological Review*, 1980.
- [19]Keith Rayner, Gretchen Kambe, and Susan A. Duffy. The Effect of Clause Wrap-Up on Eye Movements during Reading. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, Vol. 53, No. 4, pp. 1061–1080, 2000.
- [20]Masayuki Asahara. Between reading time and information structure. In *Proceedings of PACLIC*, pp. 15–24, 2017.

表 6 本実験で用いた素性.

変数名	型	記述
surprisal	実数	サブライザル
RT	実数	読み時間 (ms)
article	カテゴリカル	記事番号
screenN	整数	画面呈示順
lineN	整数	何行目か
segmentN	整数	画面右から文節目か
sentN	整数	何文目か
tokenN	整数	文内で何文節目か
length	整数	文節の文字数
freq	実数	文節の頻度 (構成するサブワードの頻度の幾何平均)
is_first	ブーリアン	行内最左要素
is_last	ブーリアン	行内最右要素
is_second_last	ブーリアン	行内右から 2 番目の要素
BOS	ブーリアン	文頭要素
EOS	ブーリアン	文末の要素
pre_EOS	ブーリアン	文末から 2 番目の要素
subj	カテゴリカル	被験者番号
WLSPLUWA	カテゴリカル	例語範疇
WLSPLUWB	カテゴリカル	意味範疇
anti_locality	整数	先行分脈における掛かり元要素数
CLMHST	ブーリアン	名詞節の節末であるか
CLMMST	ブーリアン	連体節の節末であるか
CLMFUT	ブーリアン	副詞節の節末であるか
CLMHRT	ブーリアン	並列節の節末であるか
CLMHSL	カテゴリカル	名詞節の種類 (節末文節に付与)
CLMMSL	カテゴリカル	連体節の種類 (節末文節に付与)
CLMFUL	カテゴリカル	副詞節の種類 (節末文節に付与)
CLMHRL	カテゴリカル	並列節の種類 (節末文節に付与)
infostatus	カテゴリカル	書き手の立場の情報の新旧 (以下, 名詞句に付与)
definiteness	カテゴリカル	対象を読み手が同定できるか
animacy	カテゴリカル	対象が生きているか
commonness	カテゴリカル	読み手にとっての情報の新旧. 想定可能, どちらもないも可

## A 素性

本実験で用いた素性を表 6 に示す. なおヒトに対する読み時間測定の際文末に改行を加えているため, 例えば `is_first` は「文頭である」か「文が長く画面上で改行が生じた際の改行後の文節」で発火する素性である. (`is_first`, `is_last`, `is_second_last`) について, 画面上での改行を考慮しない素性が (`BOS`, `EOS`, `pre_EOS`) であり, ヒトの読み時間のモデリングの際には前者を, 言語モデルのサブライザルのモデリングの際には後者を用いた. (`lineN`, `segmentN`) と (`sentN`, `tokenN`) についても, 画面上での改行を考慮するかしないかの違いである. 画面呈示順など読み時間測定に関わる素性は, 既存研究でより詳細に記述されている [11].

## B 言語モデルの設定

アーキテクチャ (400M パラメータの Transformer, 55M パラメータの Transformer, LSTM), 学習データサイズ (1.4G, 140M, 14M サブワード) パラメータアップデート回数 (100K, 10K, 1K, 0.1K) の異なるニューラル言語モデルを学習し, PPL が最も低くなる設定とヒトの読み時間をサブライザルによってうまくモデリングできる設定の二種類を採用した. ヒトの読み時間のモデリングには, 本文の実験と同様にコーパス中の 13,148 データポイントを使用し, 以下の式によってモデリングした:

$$\begin{aligned} \log(\text{RT}) \sim & \text{surprisal} + \text{freq} + \text{length} + \text{prev\_freq} \\ & + \text{prev\_length} + \text{is\_first} + \text{is\_last} \\ & + \text{is\_second\_last} + \text{screenN} + \text{lineN} \\ & + \text{segmentN} + (1|\text{article}) + (1|\text{subj}) . \end{aligned}$$

読み時間データに対する対数尤度で言語モデルの読み時間に対するモデリング能力を評価した. 学習率や最適化アルゴリズムなどその他のパラメータについては `fairseq` のデフォルト設定を使用した.

## C 回帰式

各実験における回帰式を記述する.

### C.1 統語範疇, 意味範疇 (3.1 節)

$$\begin{aligned} \text{surprisal} \sim & \text{WLSPLUWA} + \text{WLSPLUWB} + \text{length} + \text{BOS} + \text{EOS} + \text{pre\_EOS} \\ & + \text{sentN} + \text{tokenN} + \text{anti\_locality} + (1|\text{article}) \\ & + (1|\text{seed}) . \\ \log(\text{RT}) \sim & \text{WLSPLUWA} + \text{WLSPLUWB} + \text{length} + \text{is\_first} \\ & + \text{is\_last} + \text{is\_second\_last} + \text{screenN} + \text{lineN} + \text{segmentN} \\ & + \text{anti\_locality} + (1|\text{article}) + (1|\text{subj}) . \end{aligned}$$

### C.2 統語構造 (3.2 節)

$$\begin{aligned} \text{surprisal} \sim & \text{anti\_locality} + \text{length} + \text{BOS} + \text{EOS} + \text{pre\_EOS} + \text{sentN} \\ & + \text{tokenN} + (1|\text{article}) + (1|\text{seed}) . \\ \log(\text{RT}) \sim & \text{anti\_locality} + \text{length} + \text{is\_first} + \text{is\_last} \\ & + \text{is\_second\_last} + \text{screenN} + \text{lineN} + \text{segmentN} \\ & + (1|\text{article}) + (1|\text{subj}) . \end{aligned}$$

### C.3 節境界 (3.3 節)

`wrap-up` 効果については以下の式でモデリングし, いずれの節 (`CLMHST`, `CLMMST`, `CLMFUT`, `CLMHRT`) でも回帰係数が有意に負であった.

$$\begin{aligned} \text{surprisal} \sim & \text{CLMHST} + \text{CLMMST} + \text{CLMFUT} + \text{CLMHRT} + \text{length} + \text{BOS} + \text{EOS} \\ & + \text{pre\_EOS} + \text{sentN} + \text{tokenN} + \text{anti\_locality} + (1|\text{article}) \\ & + (1|\text{seed}) . \\ \log(\text{RT}) \sim & \text{CLMHST} + \text{CLMMST} + \text{CLMFUT} + \text{CLMHRT} + \text{length} + \text{is\_first} \\ & + \text{is\_last} + \text{is\_second\_last} + \text{screenN} + \text{lineN} + \text{segmentN} \\ & + \text{anti\_locality} + (1|\text{article}) + (1|\text{subj}) . \end{aligned}$$

節の種類による違いについては, 以下の式でモデリングした:

$$\begin{aligned} \text{surprisal} \sim & \text{CLMHSL} + \text{CLMMSL} + \text{CLMFUL} + \text{CLMHRL} + \text{length} + \text{BOS} + \text{EOS} \\ & + \text{pre\_EOS} + \text{sentN} + \text{tokenN} + \text{anti\_locality} + (1|\text{article}) \\ & + (1|\text{seed}) . \\ \log(\text{RT}) \sim & \text{CLMHSL} + \text{CLMMSL} + \text{CLMFUL} + \text{CLMHRL} + \text{length} + \text{is\_first} \\ & + \text{is\_last} + \text{is\_second\_last} + \text{screenN} + \text{lineN} + \text{segmentN} \\ & + \text{anti\_locality} + (1|\text{article}) + (1|\text{subj}) . \end{aligned}$$

### C.4 情報構造 (3.4 節)

$$\begin{aligned} \text{surprisal} \sim & \text{infostatus} + \text{definiteness} + \text{animacy} + \text{commonness} \\ & + \text{length} + \text{BOS} + \text{EOS} + \text{pre\_EOS} + \text{sentN} + \text{tokenN} \\ & + \text{anti\_locality} + (1|\text{article}) + (1|\text{seed}) . \\ \log(\text{RT}) \sim & \text{infostatus} + \text{definiteness} + \text{animacy} + \text{commonness} \\ & + \text{length} + \text{is\_first} + \text{is\_last} + \text{is\_second\_last} + \text{screenN} \\ & + \text{lineN} + \text{segmentN} + \text{anti\_locality} + (1|\text{article}) + (1|\text{subj}) . \end{aligned}$$

### C.5 交互作用の解析

$$\begin{aligned} \text{reading\_effort} \sim & \text{freq} * \text{is\_lm} + \text{length} * \text{is\_lm} + \text{prev\_freq} * \text{is\_lm} \\ & + \text{prev\_length} * \text{is\_lm} + \text{BOS} * \text{is\_lm} + \text{EOS} * \text{is\_lm} \\ & + \text{pre\_EOS} * \text{is\_lm} + \text{sentN} * \text{is\_lm} + \text{tokenN} * \text{is\_lm} \\ & + \text{anti\_locality} * \text{is\_lm} + (1|\text{article}) + (1|\text{subj\_seed}) . \end{aligned}$$

$A*B$  は  $A$ ,  $B$  という固定因子と,  $A$  と  $B$  の交互作用  $A:B$  を加えることを意味する. 分析の対象とする各素性  $Z$  について, 上式の説明変数に  $Z * \text{is\_lm}$  を加え,  $Z : \text{is\_lm}$  に対する回帰係数を分析した.