GPT 系言語モデルによる国語研長単位係り受け解析

安岡孝一 (京都大学)

概要: GPT 系言語モデル上の系列ラベリングを用いて，品詞付与・係り受け解析アルゴリズムを開発した．品詞付与においては，系列ラベリングの出力部に Bellman-Ford アルゴリズムを適用することで，解析精度を向上させている．係り受け解析においては，右向きリンクは始点から終点への情報の流れに注目し，左向きリンクは逆向き (終点から始点へ) の情報の流れに注目するアルゴリズムを開発した．これらのアルゴリズムを，27 種類の GPT 系日本語モデルに適用し，国語研長単位 Universal Dependencies による解析精度評価をおこなった．

キーワード: 言語処理，品詞付与，依存文法解析，生成言語モデル

Dependency-Parsing using Japanese Causal Language Models

Koichi Yasuoka (Kyoto University)

Abstract: In this paper the author describes how to finetune sequence-labeling for part-of-speech tagging and dependency-parsing, using Japanese causal language models, such as GPT, LLaMA and Qwen. For part-of-speech tagging, we utilize Bellman-Ford algorithm to refine the sequence-labeling. For dependency-parsing, the author has developed an original sequence-labeling algorithm, in which leftward edges are treated reversely. The author investigates efficiency of the algorithms, applying them to twenty-seven Japanese causal language models.

Keywords: Natural Language Processing, Part-of-Speech Tagging, Dependency-Parsing, Causal Language Model

1 はじめに

筆者が研究代表者を務める学際大規模情報基盤共同利用・共同研究拠点公募型共同研究『単語間に区切りのない書写言語における係り受け解析エンジンの開発』(共同研究者: 山崎直樹・二階堂善弘・師茂樹・鈴木慎吾・Christian Wittern・池田巧・守岡知彦) では，これまでに多種多様な係り受け解析エンジンを開発してきた．これらの係り受け解析エンジンは，基本的に BERT 系言語モデル (RoBERTa・DeBERTa など) を用いており，GPT 系言語モデル (LLaMA・Mistral・Qwen など) は扱ってこなかった．GPT 系は，言語モデル内部の情報の流れが単方向 (文頭から文末へ) であり，双方向の BERT 系に較べて，系列ラベリング性能が劣ると考えていたからである．

しかし，これは筆者の先入観だった．GPT 系であっても，出力部に逆方向 (文末から文頭へ) の流れを 1 段だけ付加すれば，品詞付与の性能が十分に高くなる [1]．ならば，GPT 系言語モデル向けの係り受け解析アルゴリズムも，ちょっとした工夫で実現できるのではないか．本稿では，国語研長単位 Universal Dependencies [2] を題材に，GPT 系言語モデルにおける係り受け解析アルゴリズムの可能性を探る．

2 Universal Dependencies の概要

Universal Dependencies (UD) [3] は，書写言語における品詞・形態素属性・依存構造 (係り受け関係) を，言語に関わらず記述する手法である．句構造を考慮せずに係り受け関係を記述することで，言語横断性を高めており，全ての文法構造を単語間のリンクで記述するのが特徴である．

依存構造解析それ自体は，Tesnière の構造的統語論 [4] に源を発し，Мельчук の有向グラフ記述 [5] によって，一応の完成を見た手法である．その最大の特長は，いわゆる動詞中心主義によって言語横断的な記述が可能だという点にあり，Мельчук 依存文法をコンピュータ向けに洗練した UD においても，言語に関わらない記述，という特長が前面に押し出されている．UD における文法構造記述は，句構造を考慮せず，全てを単語間のリンクとして表現する．これは，Мельчук の有向グラフ記述が，単語間のリンクという形態を取っていたからであり，そういう割り切りの結果として，言語横断的な文法構造記述を可能としているのである．

UD 依存構造コーパスの交換用フォーマットとして，CoNLL-U と呼ばれるタブ区切りテキスト (文字

表 1: UD 係り受けタグ (DEPREL)

	Nominals	Clauses	Modifier Words	Function Words
Core arguments	nsubj 主語 obj 目的語 iobj 間接目的語	csubj 節主語 ccomp 節目的語 xcomp 節補語		
Non-core dependents	obl 斜格補語 vocative 呼称語 expl 形式語 dislocated 外置語	advcl 連用修飾節	advmod 連用修飾語 discourse 談話要素	aux 動詞補助成分 cop 繋辞 mark 標識
Nominal dependents	nmod 体言による連体修飾語 appos 同格 nummod 数量による修飾語	acl 連体修飾節	amod 用言による連体修飾語	det 決定詞 clf 類別詞 case 格表示

Coordination	MWE	Loose	Special	Other
conj 接続 cc 接続詞	fixed 固着 flat 並列 compound 複合	list 細目 parataxis 隣接表現	orphan 親なし goeswith 泣き別れ reparandum 言い損じ	punct 句読点 root 親 dep 未定義

コードは UTF-8) が規定されている．CoNLL-U の各行は各単語に対応しており，以下に示す 10 個のタブ区切りフィールドで構成される．

1. ID: 単語ごとに付与されたインデックスで，文ごとに 1 から始まる整数．縮約語に対しては，単語の範囲を示すのも可．
2. FORM: 語，または，句読記号．
3. LEMMA: 基底形，語幹．
4. UPOS: UD で規定された言語普遍な品詞タグ[a)]．
5. XPOS: 言語固有の品詞タグ．
6. FEATS: UD で規定された言語普遍な形態素属性のリスト．言語固有の拡張も可．
7. HEAD: 当該の単語の係り受け元 ID．係り受け元が無い場合は 0 とする．
8. DEPREL: UD で規定された言語普遍な係り受けタグ (表 1)．HEAD が 0 の場合は root とする．言語固有の拡張も可．
9. DEPS: 複数の係り受け元を持つ場合，全ての HEAD:DEPREL ペア．
10. MISC: その他のアノテーション．

ID・FORM・LEMMA は，単語そのものに関するフィールドである．UPOS・XPOS・FEATS は，単語の品詞と形態素属性に関するフィールドである．HEAD・DEPREL・DEPS は，単語の係り受けに関するフィールドである．

a) ADJ・ADP・ADV・AUX・CCONJ・DET・INTJ・NOUN・NUM・PART・PRON・PROPN・PUNCT・SCONJ・SYM・VERB・X の 17 種類．

UD における係り受け関係は，単語間の有向グラフを HEAD と DEPREL で記述する．HEAD は，その単語に入る有向枝のリンク元 ID を示しており，DEPREL は，その有向枝における係り受けタグである．ただし，HEAD が 0 の場合，その枝に入るリンク元は存在しない．リンクの本数は単語の個数に等しく，各リンクのリンク先は，全て互いに異なっている．すなわち，各単語から出るリンクは複数有り得るが，各単語に入るリンクは 1 つだけである．なお，リンクはループしない．

UD の係り受けリンクは，Мельчук 依存文法の後裔であり，いわゆる動詞中心主義である．動詞をリンク元として，主語や目的語へとリンクする．修飾関係においては，被修飾語から修飾語へとリンクする．ただし，側置詞 (前置詞や後置詞) を体言の修飾語だとみなす点 [6] が，Мельчук とは異なっている．ちなみに，コピュラ文においては，補語をリンク元として，主語へとリンクする．

国語研長単位 UD の例として,「世界中が刮目している」の CoNLL-U と，deplacy [7] による可視化を図 1 に示す．ただし，本稿のアルゴリズムでは，LEMMA・XPOS・DEPS は使用していない．

3 アルゴリズムの概要

本稿のアルゴリズムは，品詞付与と係り受け解析の 2 段階で構成される．

品詞付与は，GPT 系の系列ラベリングをそのまま使用し，トークナイザとの不一致は B- / I-ラベルで

```
# text = 世界中が刮目している
1	世界中	世界中	NOUN	名詞-普通名詞-一般	_	3	nsubj	_	SpaceAfter=No
2	が	が	ADP	助詞-格助詞	_	1	case	_	SpaceAfter=No
3	刮目し	刮目する	VERB	動詞-一般-サ行変格	_	0	root	_	SpaceAfter=No
4	ている	ている	AUX	助動詞-上一段-ア行	_	3	aux	_	SpaceAfter=No
```

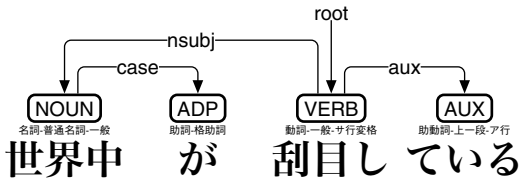

図 1: 国語研長単位 UD の CoNLL-U と deplacy による可視化

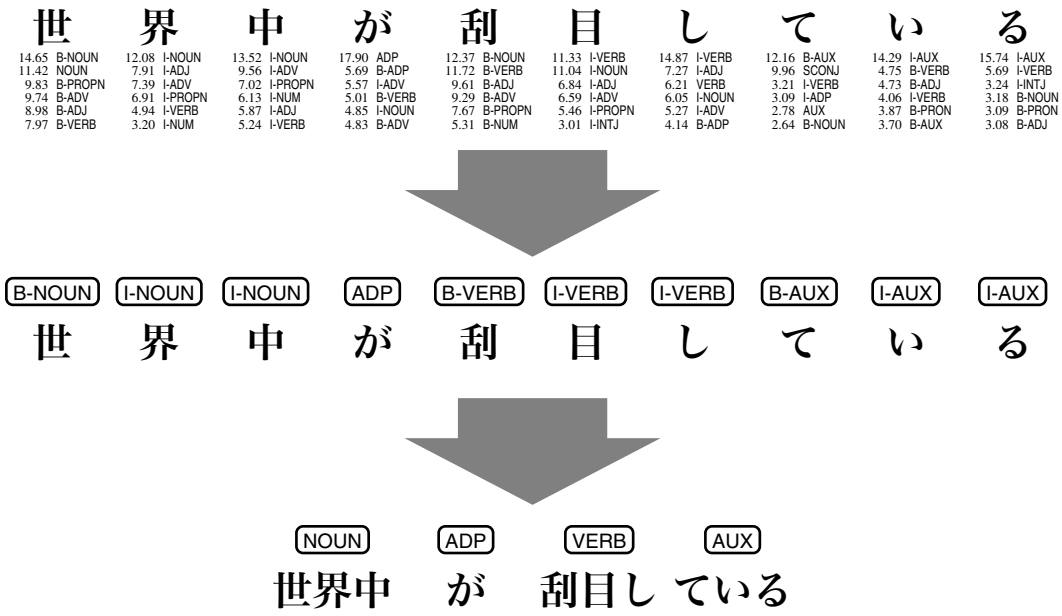

図 2: 単文字トークナイザにおける国語研長単位 UD 品詞付与と組み上げ

解消する．B- / I-ラベルの解消に際しては，Bellman-Ford アルゴリズム [8] で logits (確率のオッズ比) の合計を最大にしつつ，逆方向 (文末から文頭へ) の流れを加味する．例として「世界中が刮目している」という文を，単文字トークナイザの系列ラベリングで品詞付与する様子を，図 2 に示す．なお，組み上げた単語のベクトルは，各トークンのベクトルの合計とする．

係り受け解析は，文全体の情報を言語モデル中に留めた上で，各単語ごとに UD 係り受けリンクを抽出する．ただし，単方向 (文頭から文末へ) の情報の流れを考慮し，左向きリンクは終点の単語に注目して，始点の単語でのリンク抽出を系列ラベリングでおこなう．右向きリンクと root リンクは始点の単語に注目して，終点の単語でのリンク抽出を系列ラベリングでおこなう．図 3 の例で言えば,「世界中」では「が」への case と「刮目し」からの nsubj を抽出し,「が」では抽出をおこなわず,「刮目し」では root と「ている」への aux を抽出し,「ている」では抽出をおこなわない．最後に，これらのリンクを組み上げる形で，UD による依存文法木を得る．なお，リンクの抽出は確率的におこない，リンクに矛盾 (ループ等) が発生した場合は，logits をリンクの重みとみなした Chu-Liu-Edmonds アルゴリズム [9, 10] で，矛盾を解消する．

表 2: 本稿で使用した GPT 系言語モデルとその諸元

- https://huggingface.co/ku-nlp/gpt2-small-japanese-char
  GPT2LMHeadModel，86M パラメータ，単文字 GPT2Tokenizer，入出力幅 1024 トークン．
- https://huggingface.co/ku-nlp/gpt2-medium-japanese-char
  GPT2LMHeadModel，296M パラメータ，単文字 GPT2Tokenizer，入出力幅 1024 トークン．
- https://huggingface.co/ku-nlp/gpt2-large-japanese-char
  GPT2LMHeadModel，685M パラメータ，単文字 GPT2Tokenizer，入出力幅 1024 トークン．
- https://huggingface.co/ClassCat/gpt2-base-japanese-v2
  GPT2LMHeadModel，126M パラメータ，非バイト型 GPT2Tokenizer，入出力幅 1024 トークン．
- https://huggingface.co/llm-jp/llm-jp-1.3b-v1.0
  GPT2LMHeadModel，1.23B パラメータ，PreTokenizerFast (llm-jp)，入出力幅 2048 トークン．
- https://huggingface.co/nlp-waseda/gpt2-small-japanese-wikipedia
  GPT2LMHeadModel，106M パラメータ，ReformerTokenizer (Juman++)，入出力幅 1024 トークン．
- https://huggingface.co/nlp-waseda/gpt2-small-japanese
  GPT2LMHeadModel，106M パラメータ，ReformerTokenizer (Juman++)，入出力幅 1024 トークン．
- https://huggingface.co/nlp-waseda/gpt2-xl-japanese
  GPT2LMHeadModel，1.45B パラメータ，ReformerTokenizer (Juman++)，入出力幅 1024 トークン．
- https://huggingface.co/okazaki-lab/japanese-gpt2-medium-unidic
  GPT2LMHeadModel，322M パラメータ，BertJapaneseTokenizer (unidic-lite)，入出力幅 1024 トークン．
- https://huggingface.co/lightblue/karasu-1.1B
  LlamaForCausalLM，992M パラメータ，LlamaTokenizer，入出力幅 2048 トークン．
- https://huggingface.co/yellowback/gpt-neo-japanese-1.3B
  GPTNeoForCausalLM，1.19B パラメータ，GPT2Tokenizer，入出力幅 2048 トークン．
- https://huggingface.co/stockmark/gpt-neox-japanese-1.4b
  GPTNeoXForCausalLM，1.22B パラメータ，GPTNeoXTokenizerFast，入出力幅 1024 トークン．
- https://huggingface.co/cyberagent/open-calm-small
  GPTNeoXForCausalLM，124M パラメータ，GPTNeoXTokenizerFast，入出力幅 2048 トークン．
- https://huggingface.co/cyberagent/open-calm-medium
  GPTNeoXForCausalLM，348M パラメータ，GPTNeoXTokenizerFast，入出力幅 2048 トークン．
- https://huggingface.co/cyberagent/open-calm-large
  GPTNeoXForCausalLM，737M パラメータ，GPTNeoXTokenizerFast，入出力幅 2048 トークン．
- https://huggingface.co/cyberagent/open-calm-1b
  GPTNeoXForCausalLM，1.24B パラメータ，GPTNeoXTokenizerFast，入出力幅 2048 トークン．
- https://huggingface.co/rinna/japanese-gpt-neox-small
  GPTNeoXForCausalLM，118M パラメータ，T5Tokenizer，入出力幅 2048 トークン．
- https://huggingface.co/rinna/japanese-gpt2-xsmall
  GPT2LMHeadModel，36M パラメータ，T5Tokenizer，入出力幅 1024 トークン．
- https://huggingface.co/rinna/japanese-gpt2-small
  GPT2LMHeadModel，106M パラメータ，T5Tokenizer，入出力幅 1024 トークン．
- https://huggingface.co/rinna/japanese-gpt2-medium
  GPT2LMHeadModel，321M パラメータ，T5Tokenizer，入出力幅 1024 トークン．
- https://huggingface.co/rinna/japanese-gpt-1b
  GPT2LMHeadModel，1.21B パラメータ，T5Tokenizer，入出力幅 1024 トークン．
- https://huggingface.co/abeja/gpt2-large-japanese
  GPT2LMHeadModel，717M パラメータ，T5Tokenizer，入出力幅 1024 トークン．
- https://huggingface.co/goldfish-models/jpn_jpan_5mb
  GPT2LMHeadModel，37M パラメータ，AlbertTokenizer，入出力幅 512 トークン．
- https://huggingface.co/goldfish-models/jpn_jpan_10mb
  GPT2LMHeadModel，37M パラメータ，AlbertTokenizer，入出力幅 512 トークン．
- https://huggingface.co/goldfish-models/jpn_jpan_100mb
  GPT2LMHeadModel，119M パラメータ，AlbertTokenizer，入出力幅 512 トークン．
- https://huggingface.co/goldfish-models/jpn_jpan_1000mb
  GPT2LMHeadModel，119M パラメータ，AlbertTokenizer，入出力幅 512 トークン．
- https://huggingface.co/Kendamarron/Tokara-0.5B-v0.1
  Qwen2ForCausalLM，538M パラメータ，Qwen2Tokenizer，入出力幅 32768 トークン．

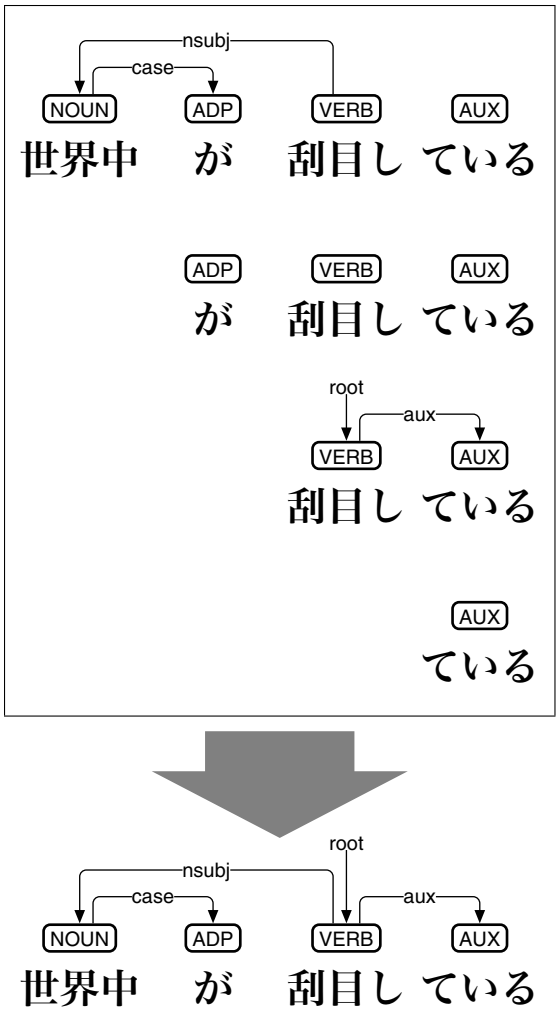

図 3: UD 係り受けリンクの抽出と組み上げ

4 評価と考察

HuggingFace ハブで公開中の GPT 系言語モデルのうち，パラメータ数が小さめの日本語モデル 27 種 (表 2) に対し，上記のアルゴリズムによるファインチューニングを国語研長単位 UD_Japanese-GSDLUW でおこない，評価 (ja_gsdluw-ud-dev.conllu による evaluation)・テスト (ja_gsdluw-ud-test.conllu による predict) をおこなった．また，[11, 12] で用いた大学入学共通テストの令和 4 年度本試験『国語』(2022 年 1 月 15 日実施) 第 1 問の問題文も，評価に加えた．いずれも評価指標は，CoNLL 2018 [13] の UPOS / LAS / MLAS [b] を用いた．結果を表 3 に示す．

京都大学大学院情報学研究科言語メディア研究室 (ku-nlp) の GPT-2 モデルが，パラメータ数に関わらず良い結果を示している．Lightblue 社の karasu-1.1B (LLaMA モデル) が，これに続いている．ku-nlp の GPT-2 モデルは単文字トークナイザを用いており，karasu-1.1B は短めのトークンを用いていることから，本稿のアルゴリズムは，短い文字数のトークンで学習した GPT 系日本語モデルに適しているようだ．他のモデルにおいては，UPOS の F 値があまり良くないことから，品詞付与の段階でしくじっており，それが係り受け解析にも悪影響を及ぼしているらしい．

たとえば yellowback/gpt-neo-japanese-1.3B は「世界中が刮目している」に対し，図 4 のように品詞付与している．これは，GPT2Tokenizer が UTF-8 のバイト列からトークンを組み上げる際に,「している」を 1 トークンにしてしまう (図 5) ためである．ならば，トークンの組み上げを途中で止めて，単文字トークナイザに近い動きをするよう改良できないだろうか．

yellowback/gpt-neo-japanese-1.3B に対し，上記の改良をおこなった上で，再ファインチューニングをおこなってみた．この改良は，GPTNeoXTokenizerFast, T5Tokenizer, AlbertTokenizer, Qwen2Tokenizer にも適用可能だったことから，それらのトークナイザを有する日本語モデルにも，同様の改良をおこなってみた．改良結果を表 4 に示す．改良をおこなった全ての日本語モデルで，UPOS の F 値が上がっており，それに合わせて LAS・MLAS も上がっている．

5 おわりに

GPT 系言語モデル上の系列ラベリングを用いて，品詞付与・係り受け解析アルゴリズムを開発した．このアルゴリズムを GPT 系日本語モデルに適用し，国語研長単位 UD での解析精度を評価した．結果として，単文字トークナイザや短めのトークンで学習したモデルに，本稿のアルゴリズムは適している．また，トークナイザを改良することで，品詞付与の精度を上げることも可能である．

ただし，本稿の結果は，BERT モデルで単文字トークナイザと Biaffine アルゴリズム [14] を用いた場合 (表 5) に，あと一歩で及んでいない．これがアルゴリズムによる差なのか，それとも GPT 系言語モデルそれ自体の限界なのか，そのあたりを明らかにしていく必要があるだろう．

本稿で作成した品詞付与・係り受け解析モデルは，`https://huggingface.co/KoichiYasuoka` で公開している．ファインチューニング用プログラムも，各

[b] 通常は LAS (Labeled Attachment Score) / MLAS (Morphology-aware Labeled Attachment Score) / BLEX (Bi-LEXical dependency score) の 3 つの評価指標を用いるが，本稿のアルゴリズムは LEMMA を使用していないため，BLEX を外し，代わりに UPOS の F 値を用いた．

表 3: GPT 系言語モデルによる国語研長単位係り受け解析の評価 (UPOS / LAS / MLAS)

	評価 (evaluation)	テスト (predict)	第 1 問【文章 I】	第 1 問【文章 II】
ku-nlp/gpt2-small-japanese-char	95.08 / 89.26 / 78.43	**94.78** / **87.85** / 77.25	**89.77** / **76.59** / 56.75	94.13 / 82.23 / **67.46**
ku-nlp/gpt2-medium-japanese-char	**95.31 / 90.98 / 81.06**	94.55 / 87.76 / **77.41**	88.66 / 75.60 / **56.95**	**94.20** / **82.89** / 67.12
ku-nlp/gpt2-large-japanese-char	94.53 / 88.87 / 78.91	94.06 / 86.91 / 76.64	87.29 / 73.36 / 53.71	90.95 / 78.45 / 61.20
ClassCat/gpt2-base-japanese-v2	89.97 / 78.07 / 66.45	88.97 / 75.13 / 63.82	84.24 / 67.32 / 45.90	89.15 / 71.38 / 52.50
llm-jp/llm-jp-1.3b-v1.0	79.29 / 61.52 / 48.95	79.41 / 61.23 / 49.47	75.65 / 52.45 / 32.72	81.47 / 59.20 / 42.95
nlp-waseda/gpt2-small-japanese-wikipedia	69.58 / 45.86 / 31.35	70.58 / 47.41 / 32.91	71.99 / 50.76 / 32.36	77.34 / 55.95 / 38.20
nlp-waseda/gpt2-small-japanese	71.05 / 48.52 / 33.43	72.06 / 49.78 / 35.52	73.34 / 51.17 / 33.65	79.41 / 59.73 / 41.64
nlp-waseda/gpt2-xl-japanese	77.14 / 57.43 / 42.96	77.04 / 56.41 / 41.96	72.70 / 49.83 / 29.92	78.35 / 56.48 / 36.36
okazaki-lab/japanese-gpt2-medium-unidic	62.03 / 33.97 / 21.20	64.28 / 36.54 / 22.70	66.72 / 40.56 / 21.69	68.92 / 42.63 / 26.61
lightblue/karasu-1.1B	93.44 / 87.73 / 76.84	93.58 / 86.31 / 75.86	86.55 / 71.93 / 51.98	91.81 / 76.84 / 60.68
yellowback/gpt-neo-japanese-1.3B	44.99 / 18.44 / 11.00	45.73 / 19.38 / 11.54	43.92 / 15.98 / 8.56	51.38 / 20.18 / 7.46
stockmark/gpt-neox-japanese-1.4b	43.30 / 17.89 / 10.34	42.65 / 17.11 / 10.13	42.13 / 15.29 / 7.56	45.35 / 15.40 / 4.81
cyberagent/open-calm-small	41.38 / 16.35 / 8.92	40.83 / 15.45 / 8.52	40.98 / 14.45 / 5.70	45.15 / 15.38 / 5.07
cyberagent/open-calm-medium	42.50 / 16.50 / 9.02	42.38 / 16.16 / 9.33	42.02 / 14.33 / 6.70	45.83 / 15.50 / 5.42
cyberagent/open-calm-large	42.43 / 16.72 / 9.33	42.25 / 16.25 / 9.57	42.32 / 15.12 / 7.39	45.89 / 15.30 / 5.54
cyberagent/open-calm-1b	42.60 / 16.68 / 9.45	42.29 / 16.51 / 9.89	41.57 / 14.55 / 7.24	45.20 / 14.68 / 4.82
rinna/japanese-gpt-neox-small	55.07 / 29.64 / 18.20	54.06 / 28.25 / 17.79	49.71 / 20.60 / 8.26	54.68 / 22.82 / 9.65
rinna/japanese-gpt2-xsmall	39.17 / 12.45 / 4.65	40.51 / 12.75 / 4.98	42.54 / 12.76 / 2.83	48.87 / 17.60 / 6.79
rinna/japanese-gpt2-small	40.60 / 14.03 / 6.00	41.76 / 14.32 / 6.04	43.64 / 13.36 / 4.26	48.47 / 17.47 / 6.17
rinna/japanese-gpt2-medium	37.94 / 12.40 / 5.28	38.65 / 12.15 / 5.34	44.64 / 13.96 / 5.15	48.16 / 17.31 / 7.26
rinna/japanese-gpt-1b	40.48 / 16.63 / 9.39	39.89 / 15.05 / 8.61	37.48 / 13.19 / 6.18	45.43 / 15.14 / 5.88
abeja/gpt2-large-japanese	40.61 / 13.41 / 5.75	42.64 / 14.09 / 6.00	44.39 / 13.74 / 4.52	49.17 / 18.22 / 5.72
goldfish-models/jpn_jpan_5mb	38.35 / 11.09 / 3.62	39.65 / 12.32 / 4.97	42.16 / 12.21 / 4.02	46.22 / 12.33 / 2.94
goldfish-models/jpn_jpan_10mb	38.55 / 11.21 / 4.02	39.23 / 11.98 / 4.55	42.21 / 12.42 / 4.17	44.58 / 12.78 / 2.34
goldfish-models/jpn_jpan_100mb	38.95 / 11.99 / 4.81	40.10 / 12.91 / 5.97	43.07 / 13.63 / 5.16	45.66 / 14.22 / 4.15
goldfish-models/jpn_jpan_1000mb	39.42 / 12.34 / 5.05	40.52 / 13.26 / 6.46	43.75 / 14.10 / 5.30	45.97 / 14.22 / 4.10
Kendamarron/Tokara-0.5B-v0.1	69.17 / 46.82 / 32.24	67.32 / 44.26 / 30.79	62.95 / 37.14 / 19.12	69.75 / 40.28 / 21.62

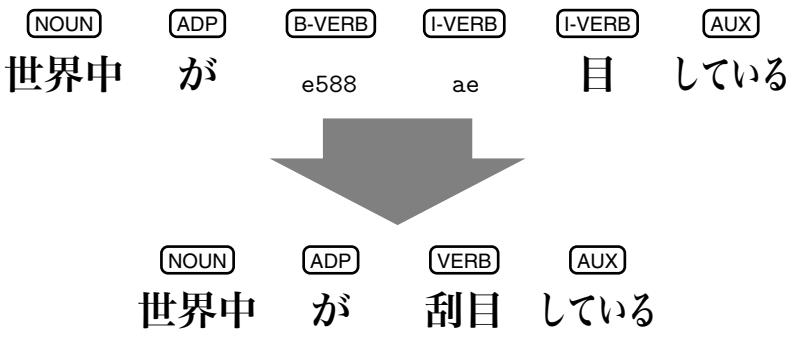

図 4: yellowback/gpt-neo-japanese-1.3B における UD 品詞付与と組み上げ

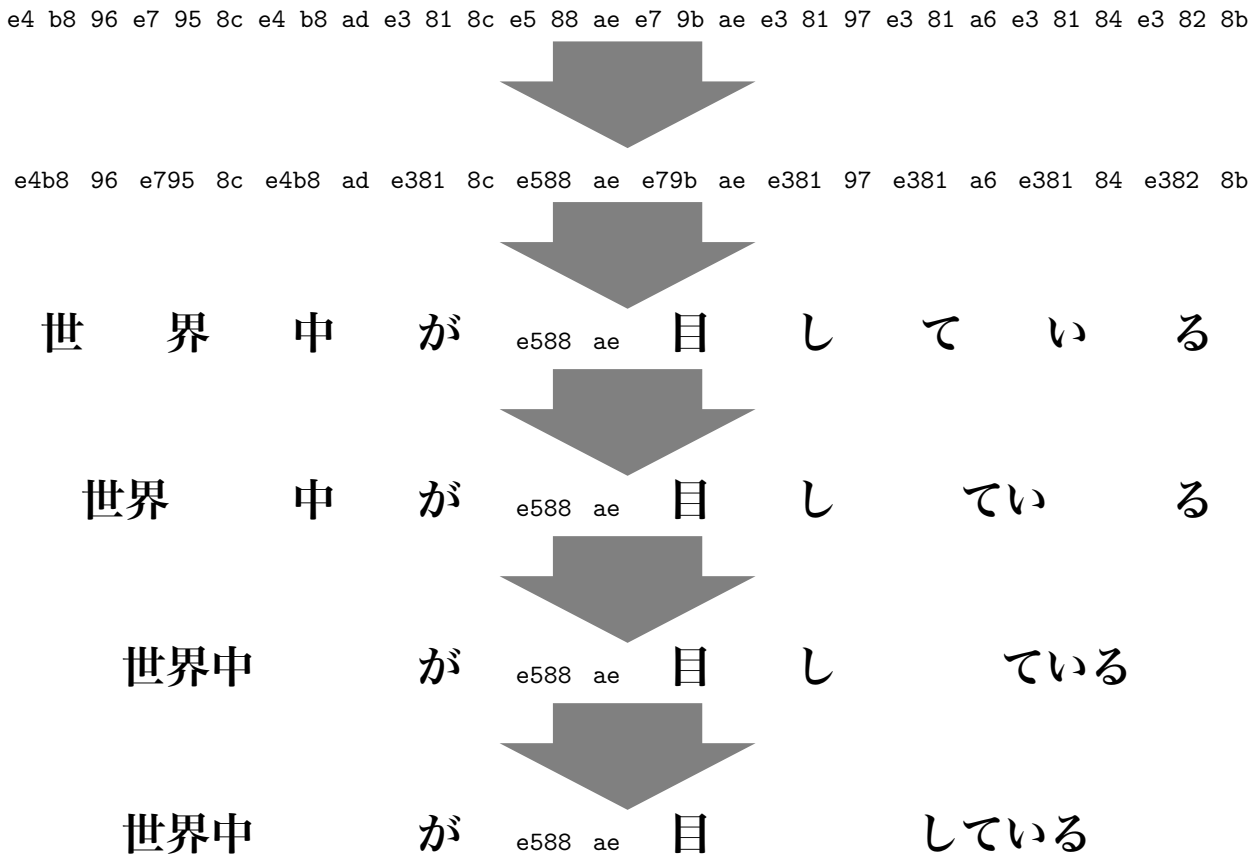

図 5: yellowback/gpt-neo-japanese-1.3B のトークナイザ動作概要

表 4: トークナイザ改良後の係り受け解析の評価 (UPOS / LAS / MLAS)

	評価 (evaluation)	テスト (predict)	第 1 問【文章 I】	第 1 問【文章 II】
yellowback/gpt-neo-japanese-1.3B	92.79 / 85.08 / 75.17	92.06 / 82.18 / 71.71	84.26 / 65.66 / 47.87	87.79 / 71.84 / 55.84
stockmark/gpt-neox-japanese-1.4b	94.86 / 88.77 / 80.02	94.22 / 86.50 / 77.98	89.10 / 76.01 / 58.47	90.78 / 76.31 / 60.68
cyberagent/open-calm-small	94.39 / 86.88 / 75.00	93.54 / 84.36 / 72.72	86.43 / 70.52 / 50.24	90.45 / 76.57 / 60.34
cyberagent/open-calm-medium	95.22 / 89.12 / 79.36	94.20 / 85.36 / 75.61	87.67 / 73.90 / 54.42	90.86 / 78.10 / 63.41
cyberagent/open-calm-large	94.86 / 88.84 / 78.82	94.44 / 86.23 / 76.74	89.12 / 75.51 / 55.93	92.05 / 78.98 / 63.97
cyberagent/open-calm-1b	95.05 / 89.07 / 80.35	94.45 / 86.53 / 78.25	87.33 / 72.55 / 54.42	92.80 / 80.11 / 65.09
rinna/japanese-gpt-neox-small	89.64 / 77.20 / 64.95	89.39 / 76.72 / 64.94	85.17 / 68.19 / 48.38	86.02 / 67.28 / 53.20
rinna/japanese-gpt2-xsmall	54.29 / 23.15 / 11.41	56.81 / 25.92 / 13.58	66.82 / 36.21 / 17.94	60.96 / 29.73 / 14.34
rinna/japanese-gpt2-small	63.20 / 33.45 / 19.73	64.08 / 34.55 / 20.70	68.58 / 38.17 / 20.79	63.86 / 33.21 / 16.21
rinna/japanese-gpt2-medium	53.06 / 22.75 / 11.68	54.82 / 24.99 / 13.45	67.12 / 38.14 / 20.95	59.62 / 29.75 / 15.04
rinna/japanese-gpt-1b	95.14 / 89.69 / 81.54	94.62 / 87.52 / 79.38	87.38 / 73.96 / 54.53	91.03 / 76.92 / 60.81
abeja/gpt2-large-japanese	54.76 / 23.62 / 12.29	57.68 / 26.75 / 14.47	68.74 / 38.52 / 21.44	61.11 / 30.74 / 15.50
goldfish-models/jpn_jpan_5mb	87.33 / 71.35 / 54.33	86.12 / 68.74 / 52.02	80.97 / 57.99 / 35.99	81.95 / 55.44 / 35.18
goldfish-models/jpn_jpan_10mb	87.74 / 71.76 / 55.06	86.56 / 70.05 / 53.25	80.31 / 58.52 / 37.23	80.89 / 55.12 / 34.80
goldfish-models/jpn_jpan_100mb	88.49 / 75.37 / 61.77	88.41 / 74.80 / 60.74	84.35 / 66.17 / 46.41	83.74 / 62.16 / 45.19
goldfish-models/jpn_jpan_1000mb	89.46 / 77.54 / 65.01	88.93 / 76.52 / 64.08	86.58 / 70.81 / 51.87	85.36 / 65.32 / 48.73
Kendamarron/Tokara-0.5B-v0.1	94.19 / 88.39 / 77.81	93.58 / 85.79 / 75.41	87.09 / 73.42 / 53.56	89.94 / 75.95 / 60.03

表 5: BERT モデルによる国語研長単位係り受け解析の現状 (UPOS / LAS / MLAS)

	評価 (evaluation)	テスト (predict)	第 1 問【文章 I】	第 1 問【文章 II】
KoichiYasuoka/bert-base-japanese-char-extended	97.39 / 92.75 / 85.19	96.99 / 91.37 / 83.51	91.88 / 81.11 / 62.70	97.76 / 89.10 / 74.01
KoichiYasuoka/bert-large-japanese-char-extended	97.90 / 93.25 / 86.43	97.67 / 92.25 / 84.77	93.97 / 83.51 / 66.67	96.55 / 86.04 / 71.70

- `https://huggingface.co/KoichiYasuoka/bert-base-japanese-char-extended` BertForMaskedLM，87M パラメータ，単文字 BertTokenizer，入出力幅 512 トークン.
- `https://huggingface.co/KoichiYasuoka/bert-large-japanese-char-extended` BertForMaskedLM，296M パラメータ，単文字 BertTokenizer，入出力幅 512 トークン.

モデルの `maker.py` で公開しているので，参考にしてほしい.

参考文献

[1] 安岡孝一: GPT 系モデルの系列ラベリングによる品詞付与, 東洋学へのコンピュータ利用, 第 38 回研究セミナー (2024 年 7 月 26 日), pp.3-10.

[2] 大村舞, 若狭絢, 浅原正幸: 国語研長単位に基づく日本語 Universal Dependencies, 自然言語処理, Vol.30, No.1 (2023 年 3 月), pp.4-29.

[3] Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, Daniel Zeman: Universal Dependencies, Computational Linguistics, Vol.47, No.2 (June 2021), pp.255-308.

[4] Lucien Tesnière: Éléments de Syntaxe Structurale, Paris: C. Klincksieck (1959).

[5] Igor A. Mel'čuk: Dependency Syntax: Theory and Practice, New York: State University of New York Press (1988).

[6] Joakim Nivre: Towards a Universal Grammar for Natural Language Processing, CICLing 2015: 16th International Conference on Intelligent Text Processing and Computational Linguistics (April 2015), pp.3-16.

[7] 安岡孝一: Universal Dependencies にもとづく多言語係り受け可視化ツール deplacy, 人文科学とコンピュータシンポジウム「じんもんこん 2020」論文集 (2020 年 12 月), pp.95-100.

[8] Richard Bellman: On a Routing Problem, Quarterly of Applied Mathematics, Vol.16, No.1 (April 1958), pp.87-90.

[9] Chu Yoeng-jin (朱永津) and Liu Tseng-hong (刘振宏): On the Shortest Arborescence of a Directed Graph, Scientia Sinica, Vol.XIV, No.10 (October 1965), pp.1396-1400.

[10] Jack Edmonds: Optimum Branchings, Journal of Research of the National Bureau of Standards—B. Mathematics and Mathematical Physics, Vol.71B, No.4 (October-December 1967), pp.233-240.

[11] 安岡孝一: Transformers と国語研長単位による日本語係り受け解析モデルの製作, 情報処理学会研究報告, Vol.2022-CH-128『人文科学とコンピュータ』, No.7 (2022 年 2 月 19 日), pp.1-8.

[12] 安岡孝一: 青空文庫 DeBERTa モデルによる国語研長単位係り受け解析, 東洋学へのコンピュータ利用, 第 35 回研究セミナー (2022 年 7 月 29 日), pp.29-43.

[13] Daniel Zeman, Jan Hajič, Martin Popel, Martin Potthast, Milan Straka, Filip Ginter, Joakim Nivre, and Slav Petrov: CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, Proceedings of the CoNLL 2018 Shared Task (October 2018), pp.1-21.

[14] Timothy Dozat, Christopher D. Manning: Deep Biaffine Attention for Neural Dependency Parsing, 5th International Conference on Learning Representations (April 2017), C25.