# 汎用言語モデル BERT を用いた 多言語テキストにおける意味現象タグ予測

伊藤 美賀 [1,a]　佐藤 七海 [1]　田上 青空 [1]
谷中 瞳 [2,1]　峯島 宏次 [1]　戸次 大介 [1]
[1] お茶の水女子大学　[2] 理化学研究所
{[a]g1720505, g1720517, g1720526,bekki}@is.ocha.ac.jp,
hitomi.yanaka@riken.jp, mineshima.koji@ocha.ac.jp

## 1　はじめに

汎用言語モデルは，文脈における語の出現確率に寄与する情報を元に，各語の表現を与えたものである．そのような情報には，品詞や構文情報，意味情報といった多様な側面が含まれており，汎用言語モデルの登場によって様々なタスクにおいて高い精度がもたらされている．特に，汎用言語モデルの一つである Bidirectional Encoder Representations from Transformers (BERT) [4] を利用することで，感情分類といった文書分類タスク，Part-of-speech tagging/形態素解析や固有表現抽出といった系列ラベリングタスクにおける精度向上が報告されている．また，大規模データによる事前学習によって，タスクの種類や言語の種類によらない横断的な言語処理が可能となった．

一方で，汎用言語モデルの言語理解能力は未だ自明ではなく，構文的理解については，汎用言語モデルが文法性を獲得できているかについて現在も研究が進められている [5, 8]．意味的理解については，質問応答や含意関係認識といった複合的なタスクにおいて評価が進められてきたが，意味の理解とは必ずしもそれらによって検証しうるものに留まるものではなく，単語レベルの評価，フレーズ・文レベルの評価，談話構造レベルの評価と，より多角的な検証が求められている段階である．

そこで本研究では，統語・意味解析情報つき多言語コーパスである Parallel Meaning Bank (PMB) [1] を用いて，各単語の意味現象タグ予測という系列ラベリングタスクで，汎用言語モデル BERT の単語レベルの意味的理解における性能を評価する．意味現象タグ (Semantic Tag) [2] とは，各単語の意味情報を分類するためのタグであり，単語の形態によって分類した品詞タグに対して，意味に特化した分類タグである．まず，BERT を用いて意味現象タグ予測モデルを構築する．構築したモデルは研究利用が可能な形式で公開予定である．また，構築したモデルを用いて，多言語で実験を行い，意味現象タグの言語間の違いを比較する．最後に，意味現象タグの文の意味合成への応用を検討する．

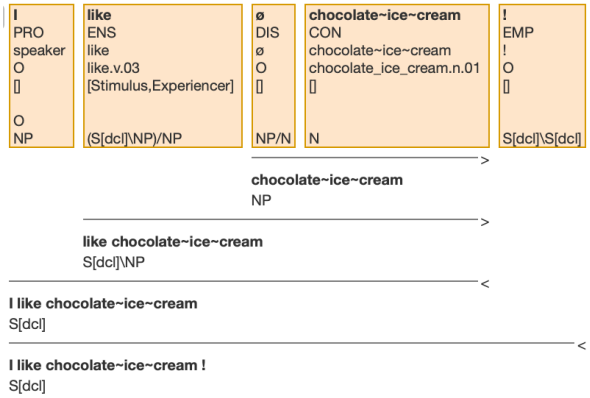


図 1: PMB のアノテーションの例

## 2　多言語統語・意味情報コーパス

### 2.1　Parallel Meaning Bank (PMB)

PMB は，多言語･多ジャンルの文について，後述する意味現象タグ [2] や，複数の意味を持つ語に対して語義を指定するための WordNet 語義タグ，組合せ範疇文法 (Combinatory Categorial Grammar, CCG) [11, 12] に基づく導出木，談話表示構造 (Discourse Representation Structure, DRS) [6] などが付与されたコーパスである．対象言語は主に英語，ドイツ語，オランダ語，イタリア語の 4ヶ国語であり，現在では，日本語，中国語と対象言語を拡張する試みが行われている．

統語・意味解析情報は，言語解析モデルによる自動アノテーションと，Bits of Wisdom (BoW) [3] と呼ばれる人手による編集によって付与される．PMB のアノテーション情報は，PMB explorer[1] という Web インターフェイスから閲覧し，編集できる．一例として，*I like chocolate ice cream!* のアノテーション情報を図 1 に示す．PMB はアノテーションの質によって，gold, silver, bronze の三種類に分類されている．gold はアノテータによって正しい情報であると判定されたデータ，silver は少なくとも一つの BoW を含むデータ，bronze は BoW を含まないデータである．

[1] https://pmb.let.rug.nl/

表 1: 意味現象タグ付与の例

| 単語 | He | carved | me | a | wooden | doll | . |
|---|---|---|---|---|---|---|---|
| 意味現象タグ | PRO | EPS | PRO | DIS | IST | CON | NIL |
| 意味現象タグの意味 | 照応代名詞 | 過去形 | 照応代名詞 | 選言, 存在 | 共通部分 | 概念 | 意味論的に空 |

表 2: 意味現象タグ（一部）

| 上位 | 下位 | 説明 | 例 |
|---|---|---|---|
| ANA | PRO | anaphoric/deictic | *he, she, I, him* |
| | DEF | definite | *the*, $lo^{IT}$, $der^{DE}$ |
| | HAS | possessive | *my, her* |
| | REF | reflexive/reciprocal | *herself, each_other* |
| UNE | CON | concept | *dog, person* |
| | ROL | role | *student, brother, victim* |
| LOG | ALT | alternative/repetition | *another, different, again* |
| | XCL | exclusive | *only, just* |
| | NIL | empty semantics | *to, of* |
| | DIS | disjunction/existential | *a, some, any, or* |
| | IMP | implication | *if, when, unless* |
| | AND | conjunction/universal | *every, and, who, any* |
| EVE | EXS | untensed simple | *to walk, destruction* |
| | ENS | present simple | *we walk, he walks* |
| | EPS | past simple | *ate, went* |
| | EXG | untensed progressive | *is running* |
| | EXT | untensed perfect | *has eaten* |

表 3: 意味現象上位タグの一覧

| 上位タグ | 説明 | 上位タグ | 説明 |
|---|---|---|---|
| ANA | Anaphoric | MOD | Modality |
| ACT | Speech Act | DSC | Discourse |
| ATT | Attribute | NAM | Named Entity |
| COM | Comparative | EVE | Events |
| UNE | Unnamed Entity | TNS | Tense & Aspect |
| DXS | Deixis | TIM | Temporal Entity |

## 2.2 意味現象タグ

意味現象タグは構成的意味論で分析対象となる言語現象に基づいて，各単語の意味情報を分類したものである．2020 年 1 月時点では 76 種類あり，さらにそれらを言語現象ごとにカテゴライズした上位タグ 14 種類からなる．表 1 に意味現象タグ付与の例，表 2 に意味現象タグの一部を，表 3 に意味現象上位タグの一覧を示す．例えば，動詞 *walk* に対して不定詞なら EXS，現在形なら ENS を付与するというように，同じ品詞の中で用法や時制の違いによって細かく区別する場合もあれば，*and* と *every* に対して同じ全称量化の意味を表す AND を付与するというように，品詞をまたがって同じタグを付与する場合もある．意味現象タグは個別の言語に依存しないため，意味現象タグを用いることで，特定の言語の構文や形態にとらわれずに，多言語間で文の意味解釈，比較を行うことができる．

各言語における意味現象タグごとの頻度と傾向を調査した結果を図 2 に示す．各言語の特徴が顕著に現れたのは，定表現 (definite) を意味する DEF という，英語の定冠詞 *the* などに付与される意味現象タグである．DEF の出現頻度を言語間で比較すると，イタリア語では多く，日本語では極端に少ない．これは英語では (1) に示すように所有格 *my*（意味現象タグは HAS）は定冠詞とともには使われないのに対して，イタリア語では (2) に示すように所有格 *mio* とは別に定冠詞 *il* がつき，日本語では (3) に示すように定冠詞 *the* に相当する単語が存在しないからである．

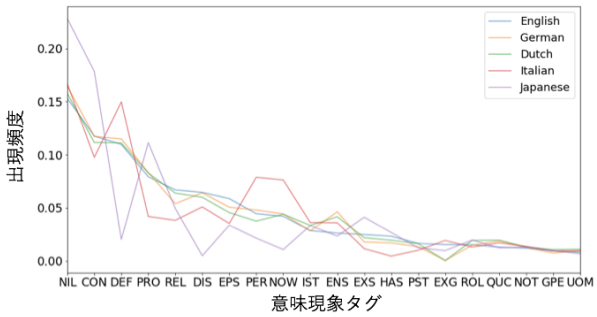


図 2: 各言語ごとの意味現象タグの分析結果

(1) 英: I'm worried about my weight
HAS CON

(2) 伊: Io sono preoccupata per il mio peso
DEF HAS CON

(3) 日: 体重 が 気 に かかる
CON

# 3 意味現象タグ予測モデルの構築

## 3.1 実験設定

本稿では，大規模テキストによる事前学習で学習した汎用言語モデルを用いることで，テキスト中の各単語の意味情報を高精度に予測できるかについて分析を行うため，汎用言語モデルの一つである BERT [4] を用いて意味現象タグ予測モデルを構築した．BERT は双方向 Transformer をベースとした汎用言語モデルであり，大規模コーパスで事前学習を行い，その後タスクに応じたファインチューニングを行うことで，様々なタスクに応用できるモデルである．また，BERT には，英語のコーパスで事前学習を行った単言語モデルと，多言語のコーパスで事前学習を行った多言語モデルがある．本稿では，pytorch の事前学習済み BERT[2]の単言語モデルと多言語モデルの両方について，英語，ド

[2]https://github.com/huggingface/pytorch-pretrained-bert

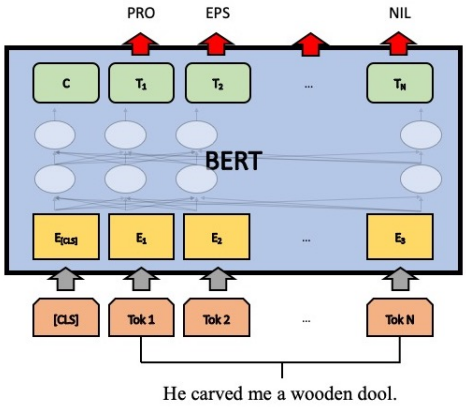


図 3: BERT を用いた意味現象タグ付与

イツ語，オランダ語，イタリア語， 日本語の 5 言語における意味現象タグ予測タスクでファインチューニングを行い，精度比較を行った．図 3 に BERT による意味現象タグ付与の例を示す．ファインチューニングのパラメータ更新は Adam[7]，ロス関数はクロスエントロピーを用いた．意味現象タグの予測精度は 10 分割交差検証の正答率で評価した．

また，ベースラインとして，Bidirectional LSTM (BiLSTM) による系列ラベリングモデルとの比較を行った．さらに，データのアノテーションの質と予測精度との相関を分析するため，PMB の gold データ，silver データ，gold+silver データの 3 種類で精度比較を行った．なお，日本語に関しては，実験時は gold データが存在しなかった[3]ため，bronze データを用いた．

## 3.2 実験結果

意味現象タグ予測モデルの評価結果を表 4 に示す．

**モデル間の比較** gold データでベースラインモデルと BERT 単言語・多言語モデルとを比較すると，日本語以外の全言語で BERT の方が精度が高かった．日本語は自動付与されたデータを用いており，UNK（不明）タグの割合が多いため，BERT の事前学習の効果がなかったと考えられる．BERT 単言語モデルと多言語モデルとを比較すると，全言語において，gold データと silver データでは，単言語モデルよりも，多言語モデルの方が精度が高かった．gold+silver データの場合もほぼ同様の結果だが，英語だけ精度が飽和状態にあり，わずかに単言語モデルでの精度の方が高かった．

**各言語の正答率と言語間の比較** 英語は，どのデータを用いた場合も 9 割を超える精度が得られた．ドイツ語，イタリア語は，いずれの場合も 8 割を超えていた．特にドイツ語は，全ての結果で 8 割 5 分を超えており，英語の次に安定した精度が得られた．オランダ語は，gold データで高い精度が得られたものの，silver データを用いた際は，単言語版では 8 割を切り，安定した精度が得られなかった．日本語は gold データがないため，bronze データを用いたが，単言語版でも多言語版でも 8 割前後の精度であり，5 言語の中で最も精度が低かったといえる．

**アノテーションの質（gold, silver）と正答率の関係** アノテーションの質が高い gold データ，gold+silver データ，silver データの順に高い精度が得られると予測していたが，必ずしもそのようではなかった．gold+silver データでの精度と gold データでの精度を比較すると，イタリア語のみ，BERT 単言語版において， gold+silver データでの精度の方が高かった．以上の結果から，データの質だけでなく量も精度に影響することが示唆された．

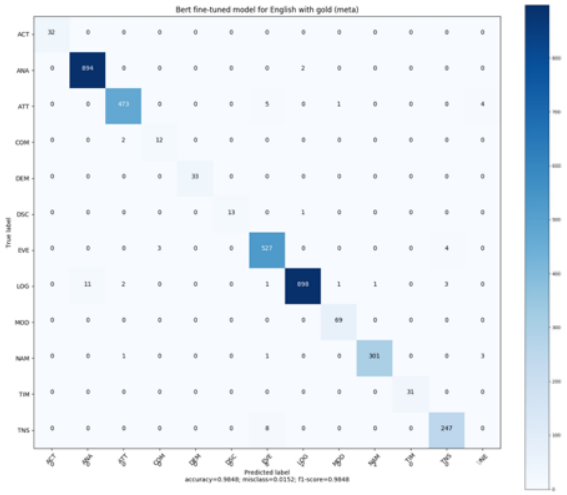


図 4: 英語の意味現象上位タグの結果（gold データ）

**タグごとの混同行列（英語）** 図 4 は，英語の意味現象上位タグの予測結果を表している．縦軸が正解ラベル，横軸が予測ラベルを表しており，色が濃いほど件数が多かったことを意味する．対角線状に，色がついた正方形が見られることから，予測ラベルの正答率は高かったと言える．また，ANA（anaphoric/照応）と LOG（logical/論理）に最も濃い色が付いていることから，この 2 つのラベルの正答率が，最も高かったといえる．

**エラー分析** 予測が正解と異なった意味現象タグを含む文を抽出し，言語ごとにエラー分析を行った．英語とドイツ語，オランダ語に関しては，CON（concept/概念名詞）を他の意味現象タグと予測するケースが多かった．イタリア語に関しては，DEF（definite/定表現）をゼロ冠詞と誤って予測するケースが多かった．以下は，オランダ語のエラーを含む文の例である．

(4) Een dollar is gelijk aan honderd dollarcent
1 ドルは 100 セントに等しい (id:63/2262)

[3]現在では専門家によるアノテーションが行われており，日本語 gold データの拡張が進められている．

| 言語 | Gold | Silver |
|---|---|---|
| EN | 13,019 | 105,813 |
| DE | 2,666 | 842 |
| IT | 1,483 | 530 |
| NL | 1,344 | 338 |
| JA | 89,560 | |

表 4: 意味現象タグの数と精度比較.

| | | EN | DE | IT | NL | JA |
|---|---|---|---|---|---|---|
| BERT 単言語 | gold | **0.98** | 0.88 | 0.84 | 0.86 | |
| | silver | 0.95 | 0.85 | 0.81 | 0.76 | 0.77 |
| | gold+silver | 0.95 | 0.87 | 0.87 | 0.78 | |
| BERT 多言語 | gold | **0.98** | **0.94** | **0.91** | **0.90** | |
| | silver | 0.95 | 0.85 | 0.85 | 0.82 | 0.81 |
| | gold+silver | 0.94 | 0.87 | 0.87 | 0.84 | |
| BiLSTM (baseline) | gold | 0.94 | 0.84 | 0.78 | 0.71 | **0.82** |

※ EN は英語，DE はドイツ語，IT はイタリア語，NL はオランダ語，JA は日本語を表す.

この文の *dollar* をモデルは UOM（unit of measurement/測定単位）と予測したが，この文脈において正しいタグは CON（concept/概念名詞）である.

## 4 意味合成と推論への応用

意味現象タグの応用例の一つに，意味現象タグを用いて文から論理式に基づく意味表示を導出することが挙げられる．本研究では，形容詞の意味表示に焦点を当てて，文をイベント意味論に基づく高階述語論理式に変換し，自動推論を行う統合的システム ccg2lambda [9] の改良を試みた．構成的意味論の観点から，形容詞には，intersective な形容詞（外延的形容詞）や subsective な形容詞（内包的形容詞）などいくつかのタイプがあり，適切な意味合成を与えるには，タイプごとに異なる語彙項目を記述する必要がある [10]．しかし，形容詞の表層形ごとに語彙項目を記述すると，語彙項目の数が膨大になる上に，未登録の語彙は導出できないといった問題がある．そこで，意味現象タグ IST (intersective), SST (subsective) を用いて，次の 2 つの語彙項目を ccg2lambda に追加した.

- category : $N/N$
  semantics : $\lambda E.\lambda F.\lambda x.(F(x) \wedge E(x))$
  stag : IST
- category : $N/N$
  semantics : $\lambda E.\lambda F.\lambda x.(E(F,x) \wedge F(x))$
  stag : SST

この改良によって，統一的に形容詞のタイプを区別して文の意味合成と推論を行うことができ，次の例 (5)(6) のような含意関係を正しく判定できるようになった.

(5) a. John is a [IST vegetarian] student
vegetarian(john) ∧ student(john)
b. John is a vegetarian 【含意】

(6) a. John is a [SST skillful] surgeon
skillful(surgeon, john) ∧ surgeon(john)
b. John is skillful 【非含意】

## 5 おわりに

汎用言語モデル BERT は Part-of-speech tagging/形態素解析において高精度を達成しており，これを PMB の意味現象タグ予測に応用する試みは自然といえる．また，BERT の構成は個別言語に依存する部分が少ないため，意味現象タグの普遍性を謳う PMB との組み合わせもまた自然といえる．本研究の実験結果は，BERT の汎用性と PMB の意味現象タグの普遍性の恩恵に与るには，データの数と質がともに重要であることを示している．しかし，ファインチューニングのデータ数が少ないときには，BERT 多言語モデル，すなわち多言語での事前学習がプラスに働くことも示唆されている．今後はこの問題への更なる考察とともに，意味現象タグの意味合成や推論への応用を進めていく.

**謝辞** 本研究の一部は，JST AIP-PRISM JPMJCR18Y1，および JSPS 科研費 JP18H03284 の助成を受けたものである.

## 参考文献

[1] Lasha Abzianidze, Johannes Bjerva, Kilian Evang, Hessel Haagsma, Rik van Noord, Pierre Ludmann, Duc-Duy Nguyen, and Johan Bos. The Parallel Meaning Bank: Towards a multilingual corpus of translations annotated with compositional meaning representations. In *Proc. of EACL-2017*, pp. 242–247, 2017.

[2] Lasha Abzianidze and Johan Bos. Towards universal semantic tagging. In *Proc. of IWCS-2017*, pp. 1–6, 2017.

[3] Valerio Basile, Johan Bos, Kilian Evang, and Noortje Venhuizen. A platform for collaborative semantic annotation. In *Proc. of EACL-2012*, pp. 92–96, 2012.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proc.of NAACL-HLT2019*, pp. 4171–4186, 2019.

[5] Yoav Goldberg. Assessing BERT's syntactic abilities, 2019.

[6] Hans Kamp and Uwe Reyle. *From Discourse to Logic*. Dordrecht: Kluwer Academic Publishers, 1993.

[7] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, Vol. abs/1412.6980, .

[8] Yongjie Lin, Yi Chern Tan, and Robert Frank. Open sesame: Getting inside BERT's linguistic knowledge. In *Proc. of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pp. 241–253, 2019.

[9] Koji Mineshima, Pascual Martínez-Gómez, Yusuke Miyao, and Daisuke Bekki. Higher-order logical inference with compositional semantics. In *In Proc.of EMNLP-2015*, pp. 2055–2061, 2015.

[10] Barbara Partee. Lexical semantics and compositionality. *An Invitation to Cognitive Science: Language*, Vol. 1, pp. 311–360, 1995.

[11] Mark Steedman. *The Syntactic Process*. MIT Press, Cambridge, Mass., 2000.

[12] 戸次大介. 日本語文法の形式理論. くろしお出版, 東京, 2010.