

LLM を用いた BabelNet と分類語彙表の語義アラインメント

川崎脩吾¹ 古宮嘉那子¹ 加藤祥² 浅原正幸³

¹ 東京農工大学 ² 北海道大学 ³ 国立国語研究所

s220144r@st.go.tuat.ac.jp, kkomiya@go.tuat.ac.jp
katosachi@let.hokudai.ac.jp, masayu-a@ninjal.ac.jp

概要

日本語の辞書やコーパスなどの言語資源は整備が進む一方で、語彙を上位・下位関係に基づく階層として整理し、検索・推論・語彙獲得などに利用できる形で提供する試みは十分ではない。本研究は、日本語語彙資源である分類語彙表を多言語概念辞書 BabelNet に接続し、BabelNet の概念階層を分類語彙表へ移植することで、大規模語彙に効率よく階層構造を付与することを目指す。ChatGPT を用いて分類語彙表の各レコードに BabelNet の概念単位 synset を複数許して対応付けを行い、人手で作成された正解データと比較評価する。さらに、対応付けが難しいレコードに対して言い換え語生成による補完を試み、対応付け可能範囲の拡張と限界を議論する。

1 はじめに

日本語には分類語彙表を代表としていくつかの概念辞書が存在するが、分類語彙表には上位・下位概念の表記がない。また、上位・下位概念のある概念辞書として WordNet が有名であるが、日本語 WordNet は機械翻訳を利用して作成されていることに起因する不整合が指摘されている。そのため本研究では、分類語彙表を多言語の WordNet である BabelNet に接続し、BabelNet の概念階層を分類語彙表へ移植することで、大規模語彙に効率よく階層構造を付与することを目指す。

分類語彙表を BabelNet に接続するため、本研究では分類語彙表の概念単位であるレコードと、BabelNet の概念単位である synset を対応付ける。両資源には概念粒度の差があるため、単一レコードに複数 synset の同値対応を許して同値 synset を推定する。手順として、まず各レコードの見出し語を用いて BabelNet を検索し、候補 synset 集合を生成する。次に、レコード情報と候補 synset の情報を入力と

して ChatGPT に候補ごとの同値関係を判定させる。さらに、候補 synset が得られなかったレコードに対しては、ChatGPT に見出し語を表す日英の言い換えを生成させ、その語で再検索して候補 synset を追加するフォールバックを適用し、同様の判定を再実行して対応可能範囲の改善量と限界を検証する。

人手で作成した Gold データとの比較により、同値判定で F1=0.850 を得た。また、BabelNet による分類語彙表の網羅率は約 65% と推定され、階層付与に必要な同値または上位概念の synset が得られる割合は 85% 以上と見込まれる。

本研究の貢献は以下である。

1. 大規模語彙資源に階層構造を付与する枠組みの提案
2. BabelNet による分類語彙表の網羅率の検証

2 関連研究

Miller は、語の意味を同義語集合と上位下位関係で整理した英語語彙資源 WordNet を提案した [2]。日本語についても、Princeton WordNet の synset に日本語語彙を付与した Japanese WordNet が整備された [3] が、公式に約 5% の不整合が報告されている [4]。

Navigli と Ponzetto は、WordNet と Wikipedia 等の異種資源を統合し、多言語にまたがる概念 (synset) と意味関係を提供する BabelNet を構築した [6, 7]。一方、分類語彙表は日本語語彙を体系的に分類した大規模資源であるが [1]、概念間の上位下位関係は十分に整備されていない。本研究は、BabelNet が提供する概念階層を分類語彙表へ移植することで、日本語語彙に対する大規模な階層付与を目指す点に特徴がある。

異種語彙資源間の語義の対応付けは、文字列照合・辞書翻訳・分散表現類似度にもとづく候補検索と判定が一般的である。翻訳語獲得には JMdict の

ような辞書資源が利用され [8, 9], 候補集合からの同値判定には埋め込みモデルによる検索・ランキングが広く用いられる [10]. また, 上位下位関係の獲得は Hearst パターンに代表される表現パターン抽出として古くから研究されてきた [11]. 近年は, LLM を用いて人手に近い形でテキスト分類・注釈付けを行う試みが報告され [12], 資源整備の低コスト化が期待される. 本研究はこの方向性に基づき, 候補語義に対する同値判定を LLM による注釈として行い, 得られた対応にもとづいて分類語彙表へ階層関係を付与する.

3 データセット

本研究では, 1 節に前述したとおりデータセットとして分類語彙表および BabelNet を用いて実験を行った.

分類語彙表は日本語語彙を粗粒度の階層構造を持つ分類体系で整理した概念辞書である. それぞれのレコードは, 見出し本体と, 属するカテゴリ (部門・中項目・分類項目・分類番号・段落番号・小段落番号・語番号) の情報を持つ. 各レコードに固有の分類番号では品詞を示す 1 桁の数字と, 分類を示す 4 桁の数字が小数点で連結されている. 付録の表 7 にレコードのサンプル, 図 1 に分類番号と分類体系の関係を示す. 本研究では, 小段落番号までが一致する語の集合を, 同一領域に属する語集合 (近縁語) と定義する.

BabelNet は, WordNet や Wikipedia 等を統合した多言語概念辞書である. 概念単位の synset は, 見出し語 (lemma), 語釈文 (gloss), カテゴリー (category) を持つ. 多言語概念辞書であるため複数言語の情報を同一 synset が持つことも多い. また, 上位・下位関係を含む synset 間の関係も定義されている. ただし, 情報が欠落している synset もある. 付録の表 8 に synset のサンプルを示す. 見出し語検索により, 同じ見出し語を持つ複数の synset が得られるため, これらの中から分類語彙表のレコードと同値のものを探す.

4 BabelNet と分類語彙表の語義アラインメント

本研究では, 分類語彙表の各レコードを, 多言語概念辞書 BabelNet の概念単位である synset に対応付ける. そのために, レコードごとに候補 synset を収集し, それぞれの候補につき該当レコードとの同値関係を分類する.

表 1 候補 synset によるレコードの網羅率 (N = 9,452)

内訳	件数	割合 (%)
少なくとも 1 件の候補あり	7,783	82.3
候補なし (ゼロ候補)	1,669	17.7
両系列で候補あり (JA & JMdict)	5,592	59.2
日本語検索のみ (JA only)	846	8.95
JMdict 英訳検索のみ (JMdict only)	1,345	14.2

表 2 レコード当たりの候補数の分布

系列	平均	中央値	Q1	Q3	P90	P95	最大
和集合	11.44	3	1	9	25	45	688
日本語検索	1.61	1	0	2	4	5	24
JMdict 英訳	10.73	2	0	8	24	44	688

4.1 候補 synset の収集

候補 synset 集合は, BabelNet の単語検索により得られる synset で構成される. 具体的には, 次の 2 系列の検索結果の和集合を候補集合とする:

1. **日本語検索**: 見出し語本体で BabelNet を検索し, 得られた synset を収集する.
2. **JMdict 英訳検索**: 見出し語本体を JMdict により英訳し, 得られた英語表現で BabelNet を検索し, synset を収集する.

分類番号 1.5 に属する各レコードに対し, 日本語検索と JMdict 英訳検索を行い, 候補の到達率と系列間の補完性を分析する.

表 1 に候補取得の内訳を示す. 全体の 82.3% で少なくとも 1 件以上の候補が得られ, 17.7% はゼロ候補であった. 両系列ヒットは 59.2% である一方, JMdict 英訳検索のみで候補が得られるレコード (14.23%) や, 日本語検索のみで候補が得られるレコード (8.95%) も存在し, 2 系列は互いに補完的である.

表 2 に候補数分布を示す. 中央値は 3 (四分位範囲 1-9) で多くは少数候補だった. さらに, 候補が得られたレコードに対して, 系列間重複率を計算すると, 中央値 0.083, 平均 0.224 であり, 両系列は同一候補を返す割合が必ずしも高くないことが分かった.

4.2 候補 synset のラベリング

各レコードに対して, 候補 synset との関係性をラベリングする. 候補 synset に与えるラベルを同値 (EQUAL) に加えて上位 (HYPERNYM)・下位 (HYPONYM)・無関係 (NONE) とする. EQUAL 以外の区別は, 誤り要因の分析と境界事例の把握のために導入した. 分類語彙表と BabelNet の粒度に不整

合があるため、単一のレコードが複数の同値 synset を持つことを許す。

5 Gold データセット

Gold は、候補 synset に対する関係ラベルを人手で付与したものであり、LLM による判定の定量評価、ならびに誤り類型化の根拠として用いる。以下では、Gold の設計と対象レコードの選定方法、アノテーションの判断基準、および Gold 統計と事例を順に述べる。

5.1 Gold の設計

対象母集団 Gold の対象レコードは、評価範囲を明確化し、同一領域内での粒度差・語彙差に着目した分析を可能にするため、分類番号が 1.5 で始まる領域に限定し、候補 synset 集合が 1 件以上得られるレコードのみからサンプリングした。

対象レコードのサンプリング手順 粒度の異なるレコードを均等に含めるため、分類番号にもとづく層化抽出により Gold (計 200 件) を構成する。分類番号の小数点以下 4 桁について末尾の連続する 0 の個数を $z \in \{0, 1, 2, 3\}$ で層化し (例: 1.5501 は $z = 0$, 1.5500 は $z = 2$; 0 が多いほど粗粒度)、各層から等量をサンプリングする。ただし、ある層の件数が割当数に満たない場合は当該層を全数採用し、不足分は $z = 0$ 層から追加抽出して総数を 200 件に調整する。

5.2 アノテーションの判断基準

本研究では、レコードとその候補 synset の集合をアノテーション単位とし、各候補 synset に対して単一のラベルを付与する。参照情報は、分類語彙表側 (見出し語・分類情報・近縁語)、BabelNet 側 (lemma・gloss・category・上位概念) である。

ラベル集合

- EQUAL: レコードと synset が同一概念 (同義)。
- HYPERNYM: synset がレコードの上位概念。
- HYPONYM: synset がレコードの下位概念。
- NONE: 上記いずれでもない。

判断基準

1. デフォルトは NONE: 関係ラベルはその関係が明確な場合にのみ付与する。
2. gloss を優先: synset の意味は gloss を一次証拠とし、lemma や上位概念は補助とする。
3. EQUAL の条件: レコードの語義と synset の語義

表 3 Gold データの統計

項目	件数	割合
全体	200	100%
EQUAL を含む	158	79%
HYPERNYM を含む	26	13%
EQUAL/HYPERNYM の両方を含む	18	9%
EQUAL/HYPERNYM を含まない	34	17%

が一致し、かつ粒度差がないときに付与する。

4. 上位下位の条件: HYPERNYM/HYPONYM は「A は B の一種」で自然かつ一般に真である場合に限る。部分全体・用途・材料など、上位概念以外の関係が妥当な場合は NONE とする。
5. 情報欠落時: lemma が専門的で狭義・一義のかつ粒度一致なら EQUAL を許す。

5.3 Gold 統計と事例

Gold は 200 レコード、候補 synset は計 3,050 件である。レコード単位のラベル出現割合を表 3 に示す。候補 synset が得られたレコードのうち、79% (158 件) が 1 つ以上の同値 synset と対応した。複数の同値 synset と対応したレコードは 20% (40 件) であった。

6 実験

本節では、ChatGPT によりレコードと候補 synset をラベル付けする。推論対象は BabelNet 全体ではなく事前構築した候補集合に限定する。これにより計算量を抑えられる一方、候補生成の欠落が上限性能を決めるため、後段で候補 synset のレコードに対する網羅率と誤り要因を分析する。

6.1 プロンプト設計

基本方針は 5.2 節の判断基準をそのまま指示として与える。また LLM の同値の synset の過剰同定を抑えるため、次の追加ガイドラインを設けた。

1. (i) EQUAL は synset の示す語義の標準的な日本語訳がレコードの見出し本体と一致する場合に限定。
2. (ii) 日本語版 WordNet が情報源の synset である可能性が高いため、gloss_ja をメインに参照。
3. (iii) 概念タイプの不一致 (光と輝度など) は NONE。

表 4 ベースライン比較 (EQUAL 検出)

手法	Prec.	Rec.	F1
baseline-1	0.67	0.62	0.64
baseline-2	0.29	0.87	0.44
ChatGPT5-mini	0.88	0.83	0.85

baseline では F1 最大は Prec. 最大と同一設定であった。

6.2 埋め込み類似度ベースライン

ChatGPT による対応付けと比較するため、埋め込み表現にもとづく類似度ランキングをベースラインとして用いる。各レコードと候補 synset 集合を、所定のテンプレート (付録表 10) によりテキスト表現へ変換し、埋め込みモデルでベクトル化する。埋め込みモデルには、汎用テキスト埋め込みが検索ベンチマークで高性能であると報告された E5 を用いた [?]. その後、コサイン類似度により降順に順位付けする。

評価は Gold に対し、上位 k 件を EQUAL, それ以外を NONE とみなす 2 値設定で行った。

6.3 評価結果 (ベースライン比較)

Gold (200 レコード, $n = 3050$ ペア) に対し、ChatGPT5-mini (API) とベースラインを比較した。表 4 より、埋め込みは k を増やすと Recall は向上するが Precision が低下し、F1 は top-1 で 0.64, top-5 で 0.44 であった。これに対し ChatGPT5-mini は 2 値分類の EQUAL で F1=0.85 を達成し、top-1 ベースラインに対して F1 を +0.21 (相対+33%) 改善した¹⁾。

6.4 誤り類型と考察

全 3050 ペア中の誤りは 134 件 (4.4%) であった。主要因は、Gold が NONE にもかかわらず階層ラベル (HYPERNYM/HYPONYM) を付与する上位概念の過剰同定 (51 件, 38.1%) と、Gold が EQUAL だが NONE とする同値の取りこぼし (21 件, 15.7%) である (表 5)。上位概念の過剰同定は「強い関連」を上位概念と誤認する場合に生じ、取りこぼしは候補の情報欠損・一般性の高さにより保守的に NONE へ倒れる場合に生じた。代表的な誤り例を付録表 9 に示す。

6.5 検索語の生成による候補の拡張

本節では、候補 synset が得られなかったレコードのうち、5.1 節の手順で取り出した 100 個に対し、検

1) ChatGPT5.2 でも同様の実験を行ったが、5-mini の方が優れた結果であった。

表 5 主要な誤りパターン (gold→pred)

誤りパターン	件数	誤り内割合
NONE→HYPONYM	29	21.6%
NONE→HYPERNYM	22	16.4%
EQUAL→NONE	21	15.7%
HYPONYM→NONE	14	10.4%
NONE→EQUAL	11	8.2%
EQUAL→HYPONYM	9	6.7%
その他	28	20.9%

※ HYPERNYM→HYPONYM の取り違えは 1 件のみ。

表 6 フォールバック 実験の結果

項目	レコード単位	候補単位
対象数	100	1602
EQUAL	28 (28%)	33 (2%)
HYPERNYM	81 (81%)	171 (11%)
EQUAL+HYPERNYM (共起)	17 (17%)	-
全候補 NONE	8 (8%)	-

索語を拡張する フォールバックを導入する。分類情報を手掛かりに、ChatGPT に「同一概念の別見出し」「表記ゆれ」「近接上位語 (「概念」など粒度が大きすぎる名詞を除く)」を日英それぞれ最大 3 語まで生成させ、得られた語のみで再検索して候補集合を構築する。候補は重複除去した上で、人手で 3 値ラベリング (EQUAL/HYPERNYM/NONE) を行った。100 レコードに対し候補は計 1602 件となり、92 レコードで少なくとも 1 件の関係ラベルが得られた。内訳を表 6 に示す。EQUAL 28 件、HYPERNYM 81 件、全候補 NONE 8 件となった。EQUAL が得られなかったレコードは、日本語に特有の表現かつ、WordNet や Wikipedia などでも取り上げられない慣用句だった。この結果から、BabelNet による分類語彙表の網羅率は 65% 程度と見込める。また、階層を付与するには EQUAL または HYPERNYM の synset があれば充分であるので、階層を付与できる割合はそれより高く、85% 以上と見込める。

7 おわりに

本研究では、分類語彙表に BabelNet の synset を多対多で対応付け、BabelNet の概念階層を日本語語彙へ移植する基盤を構築した。BabelNet による分類語彙表の網羅率は約 65%、同値または上位概念が得られる割合は 85% 以上と見込まれる。レコードに対する synset のアラインメント精度では F1=0.850 を達成した。誤りは上位概念の過剰同定と同値の取りこぼし为主要因であり、情報欠損への頑健化が課題である。また、言い換え語による検索語拡張で、未探索の関連概念を追加的に発見できる可能性を示した。

謝辞

本研究は JSPS 科研費 JP22K12145, JP25K00459, 及び国立国語研究所共同研究プロジェクト「アノテーションデータを用いた実証的計算心理言語学」の助成を受けたものです。また、JMdict を教えてくださった Rowan Hall Maudslay さんにこの場を借りて御礼申し上げます。

参考文献

- [1] 国立国語研究所. 2004. 分類語彙表：増補改訂版. 大日本図書.
- [2] George A. Miller. 1995. WordNet: A Lexical Database for English. **Communications of the ACM**, 38(11):39–41.
- [3] Hitoshi Isahara, Francis Bond, Kiyotaka Uchimoto, Masao Utiyama, and Kyoko Kanzaki. 2008. Development of the Japanese WordNet. In **Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)**, Marrakech, Morocco. European Language Resources Association (ELRA).
- [4] NICT Information Analysis Laboratory, National Institute of Information and Communications Technology. 日本語 WordNet (Wn-Ja) 公式ページ. In <https://bond-lab.github.io/wnja/jpn/index.html>.
- [5] Francis Bond and Ryan Foster. 2013. Linking and Extending an Open Multilingual Wordnet. In **Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pages 1352–1362, Sofia, Bulgaria. Association for Computational Linguistics.
- [6] Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. **Artificial Intelligence**, 193:217–250.
- [7] Roberto Navigli, Michele Bevilacqua, Simone Conia, Dario Montagnini, and Francesco Cecconi. 2021. Ten Years of BabelNet: A Survey. In **Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21), Survey Track**.
- [8] Jim Breen. 2004. JMdict: a Japanese-Multilingual Dictionary. In **Proceedings of the Workshop on Multilingual Linguistic Resources**, pages 65–72, Geneva, Switzerland. COLING.
- [9] Francis Bond, Hitoshi Isahara, Kyoko Kanzaki, and Kiyotaka Uchimoto. 2008. Boot-Strapping a WordNet Using Multiple Existing WordNets. In **Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)**, pages 1619–1624, Marrakech, Morocco. European Language Resources Association (ELRA).
- [10] Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, and Furu Wei. 2022. Text Embeddings by Weakly-Supervised Contrastive Pre-training. arXiv:2212.03533.
- [11] Marti A. Hearst. 1992. Automatic Acquisition of HY-PONYMs from Large Text Corpora. In **COLING 1992**

Volume 2: The 14th International Conference on Computational Linguistics.

- [12] Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. 2023. ChatGPT outperforms crowd-workers for text-annotation tasks. arXiv:2303.15056.

付録

表7 分類語彙表レコードのサンプル：小段落番号までが同一（04-01）の語集合 犬, わんわん, 子犬 を近縁語として扱う。

レコード ID	番号	部門	中項目	分類項目	分類番号	小段落番号	語番号	見出し本体
062994		自然	動物	哺乳類	1.5501	01	01	犬
062995		自然	動物	哺乳類	1.5501	01	02	わんわん
063000		自然	動物	哺乳類	1.5501	01	07	子犬

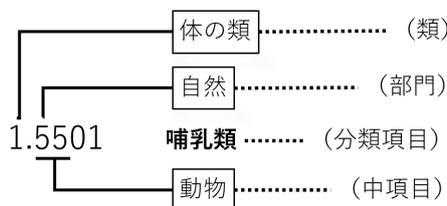


図1 分類番号と分類体系の対応（犬の例）

表8 BabelNet synset 例（酸素）：本研究で利用する主要情報のみを抽出し，冗長なメタカテゴリ等は省略した。

Field	Value (oxygen synset example)
synset_id	bn:00006912n
lemmas_ja	酸素 / オキシゲン / o
lemmas_en	oxygen / atomic_number_8 / o
gloss_ja	原子番号 8 の元素（酸素）
gloss_en	A divalent nonmetal, typically a colorless diatomic gas.
categories_en	Chemical elements ; Breathing gases ; Oxygen (他省略)
HYPERNYM_lemmas_en	chemical element ; gas ; nonmetal ; chalcogen (他省略)

表9 代表的なラベリングの誤り例

gold→pred	見出し語	候補 (lemma_en)	要因
NONE→HYPONYM	宇宙	fictional_universe	実体（物理）と設定（架空）の概念型不一致
NONE→HYPONYM	超自然	occultism	関連（領域一致）を上位概念と誤認
EQUAL→NONE	造化	universe	候補 gloss が広く，同値に踏み切れず
HYPONYM→NONE	水割り	mizuwari	説明が飲用習慣寄りで上位概念の確信が下がる

表10 生成文テンプレート（分類語彙表 / BabelNet）. 日本語情報優先. 欠損時は生成しない.

ID	使用情報	テンプレート
WLSP (H=見出し本体; C=分類項目; S=近縁語)		
W1	H	query: 見出し語=H
W2	H,C	query: 見出し語=H; 分類=C
W3	H,S	query: 見出し語=H; 近縁語=syn1 syn2 ...
W4	H,C,S	query: 見出し語=H; 分類=C; 近縁語=syn1 syn2 ...
W5	H,C,S	「H」は「C」のカテゴリに属する語であり，同じカテゴリに属する語として「syn1 syn2 ...」が挙げられる.
W6	H,C,S	query: 「H」は「C」のカテゴリに属する語であり，同じカテゴリに属する語として「syn1 syn2 ...」が挙げられる.
BabelNet (G=gloss; L=lemmas; C=category, JA/Union)		
B1	G	passage: 説明文=G
B2	G,L	passage: 説明文=G; 見出し語=lemma1 lemma2 lemma3 lemma4
B3	G,L	「lemma1」の定義文は「G」である. 同義語として「lemma2 lemma3 lemma4」が挙げられる.
B4	G,L	passage: 「lemma1」の定義文は「G」である. 同義語として「lemma2 lemma3 lemma4」が挙げられる.
B5	G,L,C	passage: 説明文="G"; 見出し語=lemma1 lemma2 lemma3 lemma4; 分類=C (cat ごとに 1 文)
各 baseline の top-k/分類語彙表テンプレート ID/BabelNet テンプレート ID/P		
baseline-1 : 1/W6/B5/JA; baseline-2 : 5/W6/B5/-		
P:JA →日本語検索で候補が取れたとき, JMdict 英訳検索による候補を無視		