

英単語学習支援に向けた語義曖昧性解消モデルの性能分析

菊地真人 按田将吾 大園忠親
名古屋工業大学大学院 情報工学系プログラム
{kikuchi,ozono}@nitech.ac.jp anda@ozlab.org

概要

語義曖昧性解消 (WSD) の技術を応用した外国語教育システムが開発され、特に語彙学習でよく利用されている。これらは、WSD の成功を前提としており、WSD の失敗は学習者の混乱を招くことにもなり得る。しかし、教育での利用を考慮した WSD モデルの性能分析が不十分である。本研究では、Transformer ベースの WSD モデル LMMS-SP の性能と、英単語に付与された語彙学習の意味での難易レベルとの関係を分析する。その結果、難しい単語ほど WSD の正解率が高い傾向にあった。また易しい単語では、WSD の正解率が3割程度しかないものがあつた。この原因として、まれにしか使わない低頻出語義を、高頻出語義と誤ることが考えられる。

1 はじめに

語義曖昧性解消 (WSD) の技術を応用した外国語教育システム [1, 2, 3] が開発され、特に語彙学習でよく利用されている。学習者は WSD を活用することで、読解文脈における単語の適切な意味 (語義) を調べる手間が省け、学習を効率化できる。しかし、Transformer [4] に基づく最先端の WSD モデルでも、WSD の正解率は7~8割にとどまる。加えて多くの教育システムは基本的に、WSD の成功を前提に設計されており、WSD の失敗は学習者の混乱を招くことになりかねない。一部の研究 [5] では、WSD の失敗を考慮した教育効果も検証されているが、WSD が常に成功するときが最も有効という結果が得られている。以上より、WSD モデルの利用が効果的な状況を、教育的側面を考慮して分析し、その性質を理解した上で用いることが重要である。それによって既存の教育システムでも、教育効果のさらなる向上を見込める可能性がある。

本研究では、Transformer ベースの WSD モデル LMMS-SP [6] の性能と、英語学習語彙表 [7, 8, 9] の難易レベルとの関係を分析する。英語学習語彙表と

は、学習者が時間や労力の観点から効率よく英単語を学べるよう、学習すべき語彙を難易レベル別に掲載したものである¹⁾。低レベルには英語を使うために不可欠な、汎用的に使う“易しい”語彙が掲載されている。高レベルには使う場面は比較的限られるが、表現力を豊かにするための“難しい”語彙が掲載されている。難易レベル毎の LMMS-SP の性能を解明することで、LMMS-SP が英語学習の下級者から上級者のどの層により有益なのかを知ることができる。さらに語義ごとの例示的なエラー分析も行う。結果として、LMMS-SP は高レベルの難しい単語ほど、WSD の正解率が高い傾向が明らかになった。また低レベルの易しい単語では、WSD の正解率が3割程度しかないものがあつた。この原因として、まれにしか使われない語義を、広く使われる語義と誤ることが考えられる。

2 関連研究

2.1 外国語教育での WSD の利活用

WSD を外国語教育に用いた最初期の研究として、Kulkarni ら [1] は英語の語彙学習システム REAP に WSD を導入し、読解文脈における英単語の語義を学習者が効率的に調べられるようにした。Rosa ら [2] は Kulkarni らの成果をベースとして、クラウドソーシングで WSD 訓練データセットを強化することで、語彙学習における WSD のさらなる有効性を示した。Degraeuwe ら [3] は、スペイン語学習のための WSD モデルとして、訓練データ収集プロセスの一部を対話的な語彙演習に統合する枠組みを提案した。これらの研究では、WSD を用いた教育システムの構築および教育効果の検証に焦点を当てている。一方で教育的側面から、既存の WSD モデルがどのような状況で有益・無益なのかは、WSD モデルの性能と関

1) 一部の語彙表では、難易レベル以外の指標で単語を区別している場合がある。本研究では分析を容易にするため、その指標をレベルに置き換えて使用している。

連付けて詳細に分析する必要がある。

2.2 WSDの性能に関する分析調査

Loureiro ら [10] は, Transformer に基づく WSD モデルの詳細な性能分析を行なった. また Kapelner ら [11] は, 人間にとって WSD を困難にする要因を明らかにした. しかし, 外国語教育での利活用を想定した, WSD モデルの分析研究は少ない. Miyata ら [12] は日本語の単語を対象に, 語義に基づく単語の曖昧度を表す公式を導出し, 単語の学習難易度と曖昧性との関係を調査した. Suzuki ら [13] は, WSD モデルを用いて英語と日本語の単語に対する曖昧度を定義し, 難易度と曖昧性の関係を再調査した. しかし, Suzuki らは WSD のために古典的な Lesk アルゴリズム [14] を用いており, 得た結論も本研究とは異なる. 彼らは難しい単語ほど曖昧性が高い, すなわち WSD が困難と主張する. 対して本研究では, Transformer による最先端の WSD モデルを用い, 難しい単語ほど WSD が容易という逆の主張をする.

3 前提知識

3.1 LMMS-SP

本研究で用いる WSD モデル LMMS-SP [6] を説明する. LMMS-SP は, 意味の埋め込み表現に基づく WSD モデルである. 本モデルは, インスタンス (WSD の対象単語) を含む文を入力とし, そのインスタンスに対する文脈埋め込みを作成する. その後, 作成した文脈埋め込みと事前訓練で得た各語義の意味埋め込みとのコサイン類似度が最大となる語義を出力する. LMMS-SP の最大の特徴は, WordNet 3.0 の収録語義を網羅した意味埋め込みを利用できる点である. WordNet [15] は電子的に利用可能な概念辞書であり, 単語の概念的な同義語, 下位語, 上位語などを語義単位で可視化することができる. 我々は英語学習支援の研究に WordNet を活用する予定であり, SOTA に近い性能を持ち WordNet の全語義を扱える LMMS-SP を選んだ.

3.2 英語学習語彙表

英語学習語彙表は, 効率的な語彙学習のために, 語彙を難易レベル別に掲載したものである¹⁾. 共通して低レベルには易しい語彙が, 高レベルには難しい語彙が掲載されている. ここでは本研究で用いる語彙表を説明する. Kilgarriff [7] は, イギリス英語

表1 北大語彙表の収録語彙に関する情報

難易レベル	単語数	学習段階
1	786	中学必修
2	1,778	高校必修
3	2,096	大学受験
4	1,520	大学基本
5	1,274	大学上級

表2 各レベルに対する平均の正解率

難易レベル	正解数	不正解数	正解率
1	1,879	806	0.700
2	1,592	529	0.751
3	594	190	0.758
4	292	67	0.813
5	118	31	0.792

の大規模コーパス British National Corpus の単語頻度をもとに BNC 語彙表を作成した. この語彙表はネイティブ英語コーパスに基づくため, 第二言語としての英語学習において, 適切な語彙が掲載されているとは限らない. 杉浦 [8] は, 高校英語の検定教科書を分析し, 使用語彙を頻度順に並べた日本人向けの語彙表を作成した. また, 英語学習への有用性を高めるため, 語彙の選定過程をより工夫した語彙表も作成されている. そのような語彙表には, 北海道大学の園田 [9] による北大語彙表, 大学英語教育学会 (JACET) の英語語彙研究会による JACET8000, アルク教育社による SVL12000 などがある. 本研究では, 各語彙表に記載の難易レベルと WSD の正解率との関係を調べ, 4.1 節と概ね同様の結論を得た. しかし本稿では紙面の都合上, 北大語彙表に絞って実験結果を掲載し, 議論する.

4 実験

実験では, 英単語の難易レベルと WSD の正解率との関係を明らかにする. また本来ならば, ある単語が持つ各語義についても, 使用頻度や学習難易度は異なるはずである. そこで, WSD を誤りやすい単語に関して, 語義単位での細かいエラー分析も行う. LMMS-SP は著者らが公開している実装をそのまま用いた²⁾. LMMS-SP で用いる Transformer は, 比較的性能が良いとされる albert-xxlarge-v2 とした. 難易レベルの指標として, 北大語彙表を用いた結果を掲載する. 北大語彙表の収録語彙に関する情報を表 1 に示す. 各レベルは学習段階で区別され, レベ

2) <https://github.com/danlou/LMMS>

表3 各レベルの英単語が持つ語義数の統計値

難易レベル	平均値	標準偏差
1	12.01	10.83
2	7.23	5.41
3	4.50	3.03
4	3.09	2.09
5	3.12	1.90

レベル1には最も平易な語彙、レベル5には最も難解な語彙が収録されている。WSDの評価データとして、Raganatoら[16]が公開したALLデータセットを用いた³⁾。このデータセットは、SensevalとSemEvalというWSDに関する各コンペティションの評価データを結合させたものである。個々のデータセットでは、高レベルの単語数が少ないものがあり、本研究での分析に不十分な場合がある。ゆえにALLデータセットを用いることにした。

4.1 難易レベルとWSD正解率との関係

北大語彙表の難易レベルごとに、WSDの正解率を平均した結果を表2に示す。どのレベルでも正解率は7割を超え、レベルが上がるほど正解率も上がる傾向にある。一般に、難解な(高レベルの)単語ほどコーパスでの出現頻度が低く、十分な訓練情報が得られない。そのため、高レベル語は正解率が低いことも予想された。しかしこの結果は、その予想とは逆の傾向を示している。低レベル語ほど様々な文脈で使われ、高レベル語ほど限定された文脈で使われることが多い。よって、高レベル語は使用機会が少ないが、持つ語義の数も少なく、WSDでは高レベル語の方が語義の特定が容易な可能性を考えた。

この仮説を検証するため、レベル毎の単語に対する語義数の平均値と標準偏差をWordNet 3.0上で調査し、表3にまとめた。この表を見ると、レベル1に属する語彙に関しては語義数の平均値と標準偏差が10を超え、語義数が多様な単語が集まっていることが分かる。全体を見ると、レベルが上がるほど語義数の平均値と標準偏差が小さくなる傾向にある。この結果は、高レベル語は語義数が少ないため、適切な語義を当てやすいという仮説を支持するものである。また表1と表2を見比べて分かるように、北大語彙表でレベル1の語彙サイズ(単語数)は小さいにもかかわらず、ALLデータセットではWSDの対象となるレベル1の単語頻度が多い。こ

3) <http://lcl.uniroma1.it/wsdeval/evaluation-data>

表4 ALLデータセットで最も出題された10語の結果

単語	難易レベル	正解数	不正解数	正解率
cancer	3	61	0	1.00
gene	4	60	0	1.00
say	1	16	26	0.38
use	1	31	10	0.76
year	1	30	6	0.83
other	1	33	0	1.00
make	1	11	20	0.35
country	1	23	8	0.74
time	1	16	14	0.53
have	1	15	15	0.50

の理由も表3から説明できる。高レベル語はそもそもWSDが不要な場合が多い。一方でレベル1の単語は、多数の語義を持つことがあり使用頻度も高いため、WSDの対象となりやすい。

4.2 語義単位のエラー分析

低レベルの基本的な語彙に対して、WSDが失敗しやすい原因を考察する。表4はALLデータセットにおいて、WSDの対象となる最頻出10語の情報を掲載したものである。大半を占める8語が、レベル1の単語であった。これらの単語の中に、正解率が高い単語と低い単語が混在する。例えば、“use”は76%の高い正解率を有するが、“say”や“make”の正解率は30%台しかない。したがって、正解率が著しく低い単語の影響で、レベル1の語彙に対する全体の正解率が低下したと考える。

そこで、正解率の低い単語“say”を例として、語義単位でのエラー分析を行った。まず、“say”の各語義に関する情報を表5に示す。なお、WordNet 3.0に存在する“say”の語義数は12であるが、この表では実験に関係する8つの語義に限って掲載した。表中のSense Keyとは、WordNetにおいて各語義に付与された固有の認識番号である。頻度とは、語義アノテーション付きコーパスSemCor 3.0における語義毎の出現頻度である。なおSemCor 3.0は、LMMS-SPの訓練データの一部であるため、高頻出の語義ほどLMMS-SPに訓練事例が多く与えられたことを意味する。品詞と説明は、語義に対してWordNet上で付与された品詞と説明である。最頻出語義(MFS)はsay%2:32:00::で、この語義は広く使用されるものであることが分かる。続いて頻出の語義はsay%2:32:01::である。残りの語義は前述の語義

表 5 単語 “say” の各語義に関する情報

Sense Key	頻度	品詞	説明
say%2:32:00::	1,867	動詞	express in words
say%2:32:01::	219	動詞	report or maintain
say%2:32:02::	4	動詞	speak, pronounce, or utter in a certain way
say%2:32:03::	27	動詞	express a supposition
say%2:32:04::	8	動詞	give instructions to or direct somebody to do something with authority
say%2:32:13::	0	動詞	state as one’s opinion or judgement; declare
say%2:32:15::	1	動詞	utter aloud
say%2:42:00::	8	動詞	have or contain a certain wording or form

表 6 単語 “say” の各語義に対する結果

予測	正解							
	2:32:00::	2:32:01::	2:32:02::	2:32:03::	2:32:04::	2:32:13::	2:32:15::	2:42:00::
2:32:00::	4/8	0	0	0	0	1	14	0
2:32:01::	1	11/14	0	0	0	2	1	0
2:32:02::	1	0	0/0	0	0	0	1	0
2:32:03::	1	0	0	1/1	0	0	0	0
2:32:04::	1	2	0	0	0/0	0	0	0
2:32:13::	0	0	0	0	0	0/3	0	0
2:32:15::	0	0	0	0	0	0	0/16	0
2:42:00::	0	1	0	0	0	0	0	0/0

2つと比較して使用頻度が低い。次に、“say”の各語義に対して、WSDを試みた結果を表6に示す。この表における対角成分は、各語義に対する正解率を表す。対角成分の分母がゼロということは、LMMS-SPがどこかのインスタンスへある語義を付与したが、実際にはその語義を正解とするインスタンスは一つもなかったことを意味する。結果を見ると、高頻出語義 say%2:32:00:: と say%2:32:01:: は正解率が高い傾向にある。一方で、まれにしか使用されない低頻出語義 (LFS) say%2:32:13:: と say%2:32:15:: に対する WSD は完全に失敗している。誤り方の傾向を確認すると、LFS を高頻出語義へと割り当ててしまうことが多いようである。

既存の WSD モデルは、MFS の割り当てに対して強いバイアスがあることが指摘されており、MFS の WSD に対しては高い性能を有するが、それ以外の LFS の性能は低いことが多い [17, 18, 19]。LMMS-SP は、WordNet の語義を網羅した意味埋め込みを活用することで、MFS fallback⁴⁾をしないことが利点であるが [20]、実験結果から LMMS-SP でも高頻出語義を過剰に割り当てる様子を確認した。基本的な単語

は多くの語義を持つが、その中でよく使われる“平易な”語義は限られる。よって、まれな語義を多く持つ単語に対して、高頻出語義の誤った割り当てが頻発したと考える。

5 おわりに

LMMS-SP の性能と英単語の難易レベルとの関係を調べた。どのレベルでも WSD の正解率は平均して7割を超え、レベルが高くなるほど正解率が高くなる (WSD が容易になる) 傾向があった。高レベル語はコーパスでの出現頻度が低いを持つ語義の数も少なく、分散表現の活用により、出現頻度が低いことによる悪影響を軽減できた可能性がある。その一方で、低レベル語は単語自体は平易でありながらも多数の語義を持ち、一部の語義しか頻繁に使われないことが WSD を難しくすることも示唆された。語義単位でのエラー分析の結果、LMMS-SP は他の WSD モデルと同様に、LFS を高頻出語義と誤ることが多いようである。LMMS-SP の教育的な応用については、難しい語彙を学びたい上級者への利用は効果的と考えられる。しかし、LFS を正確に捉えるための利用にはまだ課題が残されている。

4) 正解の語義に対する訓練データがない場合、インスタンスに MFS を割り当てること。

謝辞

本研究の一部は JSPS 科研費 JP19K12266, JP22K18006 の助成を受けたものです。

参考文献

- [1] Anagha Kulkarni, Michael Heilman, Maxine Eskenazi, and Jamie Callan. Word sense disambiguation for vocabulary learning. In **Proceedings of the 9th International Conference on Intelligent Tutoring Systems**, pp. 500–509, 2008.
- [2] Kevin Dela Rosa and Maxine Eskenazi. Impact of word sense disambiguation on ordering dictionary definitions in vocabulary learning tutors. In **Proceedings of the 24th International Florida Artificial Intelligence Research Society Conference**, pp. 507–512, 2011.
- [3] Jasper Degraeuwe and Patrick Goethals. Interactive word sense disambiguation in foreign language learning. In **Proceedings of the 11th Workshop on NLP for Computer Assisted Language Learning**, pp. 46–54, 2022.
- [4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In **Advances in Neural Information Processing Systems 30**, pp. 1–11, 2017.
- [5] Soojeong Eom, Markus Dickinson, and Rebecca Sachs. Sense-specific lexical information for reading assistance. In **Proceedings of the 7th Workshop on Building Educational Applications Using NLP**, pp. 316–325, 2012.
- [6] Daniel Loureiro, Alípio Mário Jorge, and Jose Camacho-Collados. LMMS reloaded: Transformer-based sense embeddings for disambiguation and beyond. **Artificial Intelligence**, Vol. 305, p. 103661, 2022.
- [7] Adam Kilgarriff. BNC database and word frequency lists, 1995. <https://www.kilgarriff.co.uk/bnc-readme.html>.
- [8] 杉浦千早. 高校英語教科書語彙リストの作成と使用語彙の検討. **Language Education and Technology**, Vol. 39, pp. 117–136, 2002.
- [9] 園田勝英. 大学生用英語語彙表のための基礎的研究. 北海道大学言語文化部, 1996.
- [10] Daniel Loureiro, Kiamehr Rezaee, Mohammad Taher Pilehvar, and Jose Camacho-Collados. Analysis and evaluation of language models for word sense disambiguation. **Computational Linguistics**, Vol. 47, No. 2, pp. 387–443, 2021.
- [11] Adam Kapelner, Krishna Kaliannan, H. Andrew Schwartz, Lyle Ungar, and Dean Foster. New insights from coarse word sense disambiguation in the crowd. In **Proceedings of the 24th International Conference on Computational Linguistics**, pp. 539–548, 2012.
- [12] Koki Miyata, Takahiko Suzuki, and Sachio Hirokawa. Difficulty and ambiguity of verbs: Analysis based on synsets in Japanese WordNet. In **Proceedings of the 1st International Conference on Advanced Information Technologies**, pp. 1–4, 2013.
- [13] Takahiko Suzuki, Koki Miyata, and Sachio Hirokawa. Difficulty of words and their ambiguity estimated from the result of word sense disambiguation. In **Proceedings of the 11th International Conference on Knowledge, Information and Creativity Support Systems**, pp. 1–5, 2016.
- [14] Jonas Ekedahl and Koraljka Golub. Word sense disambiguation using WordNet and the Lesk algorithm. **Projektarbeten 2004**, Vol. 17, pp. 1–6, 2004.
- [15] Christiane Fellbaum (Ed.). **WordNet: An Electronic Lexical Database**. The MIT Press, 1998.
- [16] Alessandro Raganato, Jose Camacho-Collados, and Roberto Navigli. Word sense disambiguation: A unified evaluation framework and empirical comparison. In **Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics**, pp. 99–110, 2017.
- [17] Marten Postma, Ruben Izquierdo, Eneko Agirre, German Rigau, and Piek Vossen. Addressing the MFS bias in WSD systems. In **Proceedings of the 10th International Conference on Language Resources and Evaluation**, pp. 1695–1700, 2016.
- [18] Marten Postma, Ruben Izquierdo Bevia, and Piek Vossen. More is not always better: balancing sense distributions for all-words Word Sense Disambiguation. In **Proceedings of the 26th International Conference on Computational Linguistics**, pp. 3496–3506, 2016.
- [19] Minh Le, Marten Postma, Jacopo Urbani, and Piek Vossen. A deep dive into word sense disambiguation with LSTM. In **Proceedings of the 27th International Conference on Computational Linguistics**, pp. 354–365, 2018.
- [20] Daniel Loureiro and Alípio Jorge. Language modelling makes sense: Propagating representations through WordNet for full-coverage word sense disambiguation. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 5682–5691, 2019.

A 様々な語彙表に対する難易レベルと WSD 正解率の関係

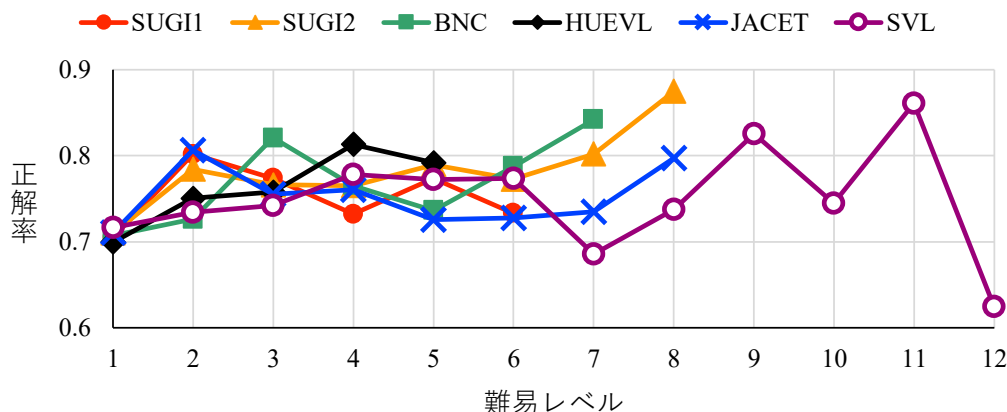


図1 様々な語彙表に対する難易レベルと WSD 正解率の関係. 各語彙表は凡例の左から, 杉浦リスト英語 I (SUGI1), 杉浦リスト英語 II (SUGI2), BNC 語彙表 (BNC), 北大語彙表 (HUEVL), JACET8000 (JACET), ALC-SVL12000 (SVL) の順で並んでいる. 語彙表ごとにレベルの粒度や掲載語数が異なるため, 同一レベルでの語彙表間の比較はできないことに注意する. SVL を除く全ての語彙表で, 正解率が最低となったのはレベル 1 であった. また, ほとんどの語彙表において高レベルの正解率が高い. なお SVL はレベル分けが細かすぎるため, レベル毎の単語数が極端に少なくなり, 高レベルの正解率が安定せず変動した可能性がある.

B 単語 “use” に関する WSD のエラー分析

表 4 において高い正解率 (76%) を示した単語 “use” に対する WSD の詳細な結果を掲載する. 表 7 は “use” の各語義に関する情報, 表 8 は各語義に対する WSD の結果である. 各表の見方は, 4.2 節で述べた表 5, 表 6 の見方と同様である. これらの表から, “use” も “say” と同様に多数の LFS を持つが, ALL データセットにおいて高頻出語義を正解とするインスタンスが多いため, 正解率が高かったことが分かる.

表 7 単語 “use” の各語義に関する情報

Sense Key	頻度	品詞	説明
use%1:04:00::	90	名詞	the act of using
use%1:07:02::	6	名詞	a particular service
use%1:22:00::	1	名詞	(economics) the utilization of economic goods to satisfy needs or in manufacturing
use%2:34:00::	8	動詞	use up, consume fully
use%2:34:01::	595	動詞	put into service; make work or employ for a particular purpose or for its inherent or natural purpose
use%2:34:02::	12	動詞	take or consume (regularly or habitually)
use%2:41:14::	0	動詞	seek or achieve an end by using to one’s advantage

表 8 単語 “use” の各語義に対する結果

予測	正解						
	1:04:00::	1:07:02::	1:22:00::	2:34:00::	2:34:01::	2:34:02::	2:41:14::
1:04:00::	3/6	0	0	0	0	0	0
1:07:02::	1	0/0	0	0	0	0	0
1:22:00::	2	0	0/0	0	0	0	0
2:34:00::	0	0	0	0/0	5	0	0
2:34:01::	0	0	0	0	28/34	0	1
2:34:02::	0	0	0	0	1	0/0	0
2:41:14::	0	0	0	0	0	0	0/1