

# 文脈を考慮した対義語穴埋め

丹羽彩奈  
東京工業大学

ayana.niwa at nlp.c.titech.ac.jp

西口佳佑

株式会社サイバーエージェント

nishiguchi\_keisuke at cyberagent.co.jp

岡崎直観

東京工業大学

okazaki at c.titech.ac.jp

## 1 はじめに

対義関係 (例: 賛成—反対) は, 与えられた文脈において意味が反対となる単語 (対義語) の関係である [1]. この対義関係を効果的に利用した修辞技法のひとつに, 似た文構造で対照的な意味を持つ単語や句を並列させる**対句**がある.

(例文) 年金に敏感な母も, 保険には鈍感です.

上記の例では, 保険への関心が低いことを, 関心の高い別の話題 (年金) と並置して強調している. このように, 対義語や対義語を用いた対句は, キャッチコピー [2] や政治演説 [3], 漢詩 [4] などでメッセージを強調するために活用されている.

従来, 与えられた単語ペアに対して対義関係を含む語彙的な意味関係を**識別する**研究が数多く行われてきた [5, 6]. 例えば, 構文パターンや意味的近さを利用して, 対義語と同義語を識別する手法が提案されている [7, 8]. しかし, その分類タスクとしての設計には, モデルが語彙を暗記してしまったり [9], 質問応答等の実タスクとの乖離があるという懸念もあった [10]. そこで, さらに挑戦的で実応用に近いタスクとして, ターゲット語 (例: 下位語) に対して特定の語彙関係にある単語や単語リスト (例: 上位語) を与えられた語彙集合から**探し出す**タスクが提案された [11]. 対義関係においても, 形容詞を否定形にする目的で対義語を**予測する**研究がある [12]. これに対して, 本研究では対義語を**文脈を考慮して予測する**新たな穴埋めタスクを提案する.

対義語の穴埋めを行うための最も単純なアプローチは, 対義語辞書を引いて, 記載されている単語を埋めることである. ところが, 穴埋めができる対義語は文脈によって変化する. 例えば, 次に示すように**母**という単一の単語に対しても, **父**, **妻**, **娘**など複

数の単語が対義語として考えられる.

(A)父は保険で楽をした。母は保険で苦労した。

(B)妻になって、値段が気になった。  
母になって、成分が気になった。

(C)娘はエステに行き、母は整体に行く。

また, 対義関係には二律相反のペア (素数—非素数) だけではなく, 相互の役割などの関係性によるもの (生徒—教師), ある基点を挟んだ相対的・対照的な概念を表すもの (明日—昨日) など多様な関係性が含まれる [13, 14, 15]. しかし, 辞書には対義語ペアのみが収録されており, その文脈情報は収録されないことが多い. また, 掲載されている対義関係の種類も限定的である. これに対し, 対義語穴埋めタスクでは, 文脈情報や多様な対義関係を考慮して対義語を予測できるモデルが求められる.

近年, モデルを事前学習およびファインチューニングするアプローチが多くタスクで高い性能を収めている [16, 17]. そこで本研究では, 代表的な事前学習モデルである BERT を利用し, 文脈を考慮した対義語穴埋めを行う. しかし, モデルをファインチューニングするための教師データ、すなわち対照的な文脈を持つ対義語ペアを含むテキストの収集は容易ではない. そこで本研究では, BERT モデルで対義語穴埋めを行うための手法として, (1) 対句データを用いたドメイン適応, (2) 対義語穴埋めに特化させるための対照的マスキング, (3) 文の対句構造を捉えさせるための対句位置エンコーディング, (4) 対義語辞書を活用した自動アノテーションによる疑似教師データ収集を提案する.

実験の結果, 提案したファインチューニング手法により, BERT モデルを文脈を踏まえた対義語穴埋めタスクへ適応させ, 正解の単語をより予測できる

ようになることを確認した。また、文脈を考慮した対義語穴埋めタスクでは複数の正解が許容されるため、唯一の正解による自動評価指標による精度は低く見えるが、対義語としての適切さや文脈における流暢さをもつ事例は85%以上を占めることから、提案手法は十分な性能を示していることを報告する。

## 2 提案手法

### 2.1 モデル

概要を図1に示したように、位置  $m$  ( $1 \leq m \leq n$ ) に [MASK] トークン (空欄) を持つ長さ  $n$  の系列  $x_1, \dots, x_n$  を与えられた時、穴埋めするトークン  $y_m$  の条件付き確率を BERT モデル [16] を用いて以下のようにモデル化する。

$$P(y_m | x_1, \dots, x_m, \dots, x_n). \quad (1)$$

入力情報の双方向の文脈で条件づけることにより、BERT モデルは [MASK] トークンの周辺の文脈を捉えられる。その BERT モデルを対句構造を持つコーパスでファインチューニングすることで、ドメイン適応により入力テキスト内の文脈と対義語ペアの組み合わせを考慮可能となることが期待される。

BERT モデルを対義語に適応させる際、少量のコーパスを活用するために二つのアプローチを取る。一つ目のアプローチとして、テキスト内に対義関係を持ち得るトークンを [MASK] トークンに置換することでファインチューニングのための教師データを作成する。例えば、「値段に始まり、味わいに終わる。」というテキストから、「[MASK] に [MASK]、味わいに終わる。」、「値段に始まり、[MASK] に [MASK]。」という二件の学習事例を得る。これらのマスクされるトークンには、対応する句に現れないものを選択する。このマスキングの方法を対照的マスキングと呼ぶ。オリジナルの BERT モデルは [MASK] トークンをランダムに作成するが、対照的マスキングは、対義語穴埋めのための教師データを選択的に作成する。

次に、二つ目のアプローチとして、BERT モデルに入力テキスト内の対句構造を考慮させるために位置エンコーディングを拡張する。まず、対句構造を構成する二つのスパン  $[i, j]$  ([MASK] トークンが含まれる) と  $[k, l]$  ( $1 \leq i < j \leq k < l \leq n$ ) を含むテキストを考える。この時、スパン  $[i, j]$  にスパン  $[k, l]$  が対応することを教えるために、対句構造に特化し

たインデックスを以下のように定義する。

$$a_t = \begin{cases} k + \left\lfloor \frac{(l-k)(t-i)}{j-i} \right\rfloor & (a \in [i, j]) \\ t & (\text{otherwise}) \end{cases} \quad (2)$$

インデックス  $a_t$  は  $t \notin [i, j]$  だったときはトークン  $x_t$  の絶対位置を、そうでない場合は対応するスパン  $[k, l]$  の絶対位置を与える。本稿では、この手法を対句位置エンコーディングと呼ぶ。今回は、オリジナルの BERT モデルで用いられている絶対位置の埋め込み表現と対句位置  $a_t$  の埋め込み表現の平均を取ったものを BERT モデルに入力する。

### 2.2 教師データ

教師データのドメインには、対句が用いられやすいキャッチコピーを選択した。既存の作品集 [18, 19, 20, 21, 22] から収集した 111,295 件のコーパスを元に、(1) 対句を含まないキャッチコピーの除外 (2) 対句スパンのアノテーションの2つのサブタスクにより、7,511 件のキャッチコピーからなる対句コーパスを作成した [23, 24]。

### 2.3 疑似教師データ

現状の教師データの学習事例数は、BERT モデルをファインチューニングするには不十分である。そこで、テキストを自動的にアノテートすることを考える。具体的には、既存の対義語辞書 [15] に掲載されている対義語ペアを含むキャッチコピーを用いる。このプロセスにより、2.2 節で説明した対句コーパスに含まれないキャッチコピーから 1,894 件を抽出した<sup>1)</sup>。各キャッチコピーに対して、対義語を片方ずつ [MASK] トークンに置換することで二件の学習事例を作成した。この外部知識を活用したデータ作成により、対義語の語彙知識を疑似教師データとして BERT モデルに学習させる。

## 3 実験および実験結果

### 3.1 実験設定

**データセット** 7,511 件の対句構造を持つキャッチコピーを学習データ、検証データ、評価データに分割した上で、[MASK] トークンを含む穴埋め事例に変換する。対句構造には二つの対照的な意味を持つ句が含まれることから、各キャッチコピーか

1) 対句コーパスに含まれないキャッチコピーは、コーパス構築の一段階目における厳しい基準でフィルタリングされたため、実際には対句を含むものが含まれている。

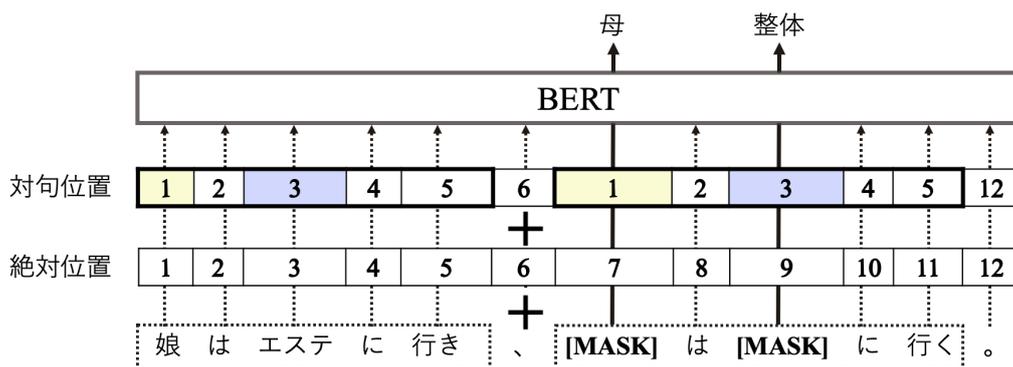


図1 提案手法の概要

ら二件の学習事例を作成できる。このようにして、11,923 件の学習データ、1,496 件の検証データ、1,247 件の評価データを作成した。この評価データの作成方法を付録 A.1 に示した。また、疑似教師データとして 3,788 件の学習データを用いる。

**ベースライン** 今回の実験では、ベースライン手法として辞書を用いた手法と、ファインチューニングなしの事前学習済み BERT モデルを採用した。辞書を用いた手法では、各空欄を埋めるべき正解単語が、対応する句に含まれる単語の対義語として辞書 [15] に含まれるかを調査する。このベースラインは、正解単語がわかっている状態で評価するため、辞書を用いた手法の性能の上限として考えられる。ファインチューニングなしの事前学習済み BERT モデルでは、対義語穴埋めに特化した学習なしに、BERT モデルが言語モデルとして文脈を考慮して対義語を予測する能力について調査する。また、3名の作業員に対して穴埋めタスクを実施した結果（以後作業員と表記）も評価する。具体的には、各空欄に対して最大 5 つの可能な回答候補を考えてもらった。3名の作業員は、本タスクの難易度を考慮し、作家やブロガーなど、作文経験の豊富な方を選んだ。なお、BERT の事前学習済みモデルには、日本語 Wikipedia で学習された公開モデル<sup>2)</sup>を用いた。

**評価** 本実験では、マスクされた位置に正しい対義語が予測されたか否かを top-1、top-10 の正解率で評価する。評価データには、[MASK] トークンがサブワード単位で分割されている事例が含まれる。しかしながら、作業員はサブワード単位で空欄を埋めることはできない。そのため、評価データのうち [MASK] トークンがサブワード分割されていない事例を含むサブセット（以後単語単位と呼ぶ）を作成し、作業員が埋めた単語とモデルが予測した単語を

比較する。また、作業員は必ずしも一つの空欄に対して 5 つの回答を思いつくとは限らないため、top-1 と top- $n$  の正解率でも評価する。この数値  $n$  は、各事例に対する回答数によって異なる。

### 3.2 実験結果

対義語穴埋めの結果を表 1 に示した。提案手法は、全評価データに対して top-1 正解率で 23.4%、top-10 正解率で 47.9%、単語単位では top-1 正解率で 30.4%、top-10 正解率で 49.1% を達成した。ファインチューニングなしの事前学習済み BERT モデルは、提案手法に比べて top-1 正解率が大きく低下している。これは、一般的なマスク言語モデルでは、対句構造の文脈を与えられていたとしても、対義語を予測するには不十分であることを示している。言い換えれば、提案手法はモデルを対句構造における単語予測タスクに適応させるのに有効であることがわかる。しかしながら、ファインチューニング無しと有りの場合で top-10・top- $n$  正解率における性能の差は top-1 に比べると小さい。これは、BERT モデルが事前学習時に文脈に基づき正解単語を正解候補として予測する能力を獲得していることを示唆している。

教師データでファインチューニングをした場合は、特に top-1 正解率が大きく向上した（全評価データで +6.1 ポイント、単語単位で +9.3 ポイント）。また、対照的マスクングにより全ての正解率が、対句位置エンコーディングにより特に top-10・top- $n$  正解率が向上した（前者は +0.8~+3.0 ポイント、後者は +1.5、+0.6 ポイント）。さらに、疑似教師データの活用により、特に top-1 の正解単語の向上を確認した（+0.8、+3.0 ポイント）。これらの提案手法が性能向上に寄与していることから、BERT モデルに対して対義語穴埋めに特化したファインチューニングを行うことが重要であるとわかる。

2) <https://github.com/cl-tohoku/bert-japanese>

表1 対義語穴埋めの正解率

	全データ		単語単位	
	正解率@1	正解率@10	正解率@1	正解率@n
辞書を用いた手法	-	-	9.6	-
BERT モデル (ファインチューニングなし)	14.5	41.1	15.7	39.1
BERT モデル (ファインチューニングあり)	20.6	45.2	25.0	44.4
- ランダムマスキング + 対照的マスキング	21.7	46.0	27.4	47.4
+ 対句位置エンコーディング	22.6	47.5	27.4	48.0
+ 疑似教師データ	<b>23.4</b>	<b>47.9</b>	<b>30.4</b>	<b>49.1</b>
作業者 (正解率が最小値となった場合)	-	-	31.5	52.3
作業者 (正解率が最大値となった場合)	-	-	34.5	59.1
作業者 (3名の回答を統合した場合)	-	-	51.8	66.6

表2 各指標を満たす100件あたりの事例数

	意味的対照性	文の自然さ・流暢さ
作業者	94	90
提案手法	88	85

辞書を用いたベースラインでは、単語単位の top-1 正解率で 9.6% という結果になった。この正解率の低さは、辞書に含まれるエントリのカバー率の低さに起因する。例えば、評価データに含まれる対義語の 39.3% しか辞書に掲載されていなかった。

表1は、穴埋めタスクを解いた時に正解率が最小・最大となった作業者の結果も示している。これを見ると、正解率は作業者ごとに揺れが見られるものの提案手法よりも高い。しかしながら、正解率は最大でも top-1 で 34.5%、top-n で 59.1% と低く、このタスクの難しさが現れている。一方で、作業者3名の回答を統合し、その中に正解単語が含まれば正解とする最も緩やかな評価を行うと、top-1 正解率は 51.8%、top-n 正解率で 66.6% となった。この top-1 正解率の大幅な向上は、評価データ内の各空欄を埋める際に複数の可能な単語が考えられることを示唆している。この正解単語が複数考えられる事象について調査するため、正解単語と一致しなかった作業者の回答と提案手法の予測結果の質について、別の作業者による主観評価を実施した。各回答・予測結果の内、意味的対照性と文としての自然さ・流暢さを満たす事例数を表2に示した。この分析では、作業者の回答と提案手法の予測単語両方が不正解だった事例から100件ランダムに抽出したものをを用いた。なお、作業者の回答は、3名の回答からランダムに1事例につき1件を選択した。この結果、作業者の回答・提案手法の予測単語ともに、85%以上の

単語が対義語として適切であることがわかった。

以上より、文脈を考慮した対義語穴埋めでは、複数通りの正解単語が存在するため、単一の正解単語を用いた自動評価（正解率）では性能が過小評価されてしまうことがわかった。しかし、提案手法による予測結果を主観評価したところ、対義語としての適切さ、文としての自然さ・流暢さも非常に高い正解率となっており、文脈を考慮した対義語を予測するには十分な性能であることがわかった。また、事例分析の結果を付録A.2に示したように、さらに性能を向上させるためには単語ペアを「何を対象として」「どのような観点で対比させるか」を明示的に与える必要があることもわかった。

## 4 おわりに

本研究では、文脈を踏まえた対義語穴埋めという、対義語の予測タスクに取り組んだ。BERTモデルを対義語穴埋めに適応させる方法として、対句データを活用したドメイン適応、対照的マスキング、対句位置エンコーディング、疑似教師データの収集方法を提案した。提案手法は、評価データにおいて、23.4%の top-1 正解率および 47.9%の top-10 正解率を達成した。これらの値は低く見えるが、空欄を埋められる正解単語は複数考えられるため、単一の正解単語に基づく自動評価では性能が過小評価されてしまう。しかし、主観評価により、提案手法による予測結果は、85%以上の事例が対義語として適切であることがわかった。本研究の結果は、語彙の意味関係を実用的な単語予測タスクとして扱う研究をさらに促進させるものである。今後は、言及対象や対比させるべき観点を考慮した対義語予測および他の意味関係へのモデルの拡張に取り組みたい。

## 参考文献

- [1] Steven Jones, M Lynne Murphy, Carita Paradis, and Caroline Willners. *Antonyms in English: Construals, constructions and canonicity*. Cambridge University Press, 2012.
- [2] Hristo Katrandjiev, Ivo Velinov, and Kalina Radova. Usage of rhetorical figures in advertising slogans. *Trakia Journal of Sciences*, Vol. 14, No. 03, pp. 267–274, 2016.
- [3] John Heritage and David Greatbatch. Generating applause: A study of rhetoric and response at party political conferences. *American journal of sociology*, Vol. 92, No. 1, pp. 110–157, 1986.
- [4] Rui Yan, Cheng-Te Li, Xiaohua Hu, and Ming Zhang. Chinese couplet generation with neural network structures. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2347–2357, 2016.
- [5] Oren Barkan, Avi Caciularu, and Ido Dagan. Within-between lexical relation classification. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 3521–3527, Online, 2020.
- [6] Vered Shwartz and Ido Dagan. Path-based vs. distributional information in recognizing lexical semantic relations. In *Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon (CogALex - V)*, pp. 24–29, Osaka, Japan, 2016.
- [7] Kim Anh Nguyen, Sabine Schulte im Walde, and Ngoc Thang Vu. Distinguishing antonyms and synonyms in a pattern-based neural network. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pp. 76–85, Valencia, Spain, 2017.
- [8] Kim Anh Nguyen, Sabine Schulte im Walde, and Ngoc Thang Vu. Integrating distributional lexical contrast into word embeddings for antonym-synonym distinction. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 454–459, Berlin, Germany, 2016.
- [9] Omer Levy, Steffen Remus, Chris Biemann, and Ido Dagan. Do supervised distributional methods really learn lexical inference relations? In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 970–976, Denver, Colorado, 2015.
- [10] José Camacho-Collados. Why we have switched from building full-fledged taxonomies to simply detecting hypernymy relations. *CoRR*, Vol. abs/1703.04178, , 2017.
- [11] Jose Camacho-Collados, Claudio Delli Bovi, Luis Espinosa-Anke, Sergio Oramas, Tommaso Pasini, Enrico Santus, Vered Shwartz, Roberto Navigli, and Horacio Sag-gion. SemEval-2018 task 9: Hypernym discovery. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pp. 712–724, New Orleans, Louisiana, 2018.
- [12] Laura Rimell, Amandla Mabona, Luana Bulat, and Douwe Kiela. Learning to negate adjectives with bilinear models. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pp. 71–78, 2017.
- [13] Thmas R. Hoffmann. *Realms of meaning: An introduction to semantics*. Longman, 1993.
- [14] Geoffrey Leech. *Semantics*. Penguin Books, 1976.
- [15] 三省堂編修所 (編). 反対語対立語辞典. 三省堂, 2017.
- [16] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, 2019.
- [17] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, Vol. 32, pp. 5753–5763. Curran Associates, Inc., 2019.
- [18] 谷山雅計. 広告コピーってこう書くんだ!読本. 株式会社宣伝会議, 2007.
- [19] 仲畑貴志. ホントのことを言うと、よく、しかられる。勝つコピーのぜんぶ. 株式会社宣伝会議, 2018.
- [20] 青田光章, 秋山 晶, 東 秀紀, ほか. 最新約コピーバイブル. 株式会社宣伝会議, 2007.
- [21] 梅田悟司. 「言葉にできる」は武器になる。日本経済新聞出版社, 2016.
- [22] 宣伝会議賞実行委員会. *SKAT.2-SKAT.17*. 株式会社宣伝会議, 2003-2018.
- [23] 丹羽 彩奈, 脇本 宏平, 西口 佳佑, 毛利 真崇, 岡崎 直観. キャッチコピーにおける対句構造の解析. 言語処理学会第 26 回年次大会 (NLP2020), 2020.
- [24] 丹羽 彩奈, 脇本 宏平, 西口 佳佑, 毛利 真崇, 岡崎 直観. 単語の対応関係を利用したスパン候補の絞り込みによるキャッチコピーの対句構造解析. 第 34 回人工知能学会全国大会 (JSAI2020), 2020.

## A 付録

### A.1 評価データの作成方法

評価データ内で空欄とする単語（列）には、2.1 節で説明した対照的マスキングの対象となった単語（列）から、以下の基準に基づき選択した。

1. 評価事例一件につき一単語（句は除外）
2. 単語の品詞は名詞、動詞-自立、形容詞-自立、形容動詞のいずれか
3. 企業情報などの外部知識がなければ解けない事例は可能な限り除外

本来は一件のキャッチコピーから前半の句と後半の句をそれぞれマスキングすることで二件の評価事例を作成できる。しかし、クラウドソーシングプロセスの簡略化のため、単語単位ではキャッチコピー一件につき評価事例一件のみを使用した。そのため、元々の評価データに比べて小規模になっている。

### A.2 事例分析

ベースライン（ファインチューニングなしのBERTモデル）、提案手法、各作業者による出力・回答例を表3に示した。

事例(A)の（出会い—別れ）のように、対比関係がわかりやすい単語が存在する場合は、提案手法も作業者も正解単語を第一候補として出力することができる。また、予測すべき単語の周辺のみを考慮しているベースラインの出力と比べると、提案手法は「別れの曲—出会いの曲」というテキスト内の大域的な構造に着目できていることがわかる。

事例(B)は、正解単語は予測・回答されていないものの、対義語として機能している。この問題では、「地球」という単語に対して、規模の大きさや身近さの度合いで対比できる単語を出力する。その点、提案手法の予測結果・作業者が作成した回答ともに自分の周りにある物や人を中心とした単語が並んでおり、どれも意味的対照性と文としての自然さを満たしている。しかし、正解単語はそれらの観点に加えて「物理的・精神的な対比」を持つ「心」となっている。このような事例に対応するためは、「二つの単語をどのような観点で対比させるのか」を明確にすることが必要であると考えられる。

また、人間にとっても正解することが難しい他の

表3 対義語穴埋めタスクの出力・回答例

(A) 別れの曲だったのに、[MASK]の曲になった。	
正解	出会い
ベースライン	別れ, 最後, 今, 人生, 卒業
提案手法	出会い, 憧れ, 人生, 最高, 始まり
作業者1	出会い, 再会, 初恋, 永遠, 永久
作業者2	出会い, 始まり, スタート, 開始, 邂逅
作業者3	出会い, 出逢い
(B) 地球の環境より、まず[MASK]の環境。	
正解	心
ベースライン	宇宙, 水, 地球, 太陽, 植物
提案手法	家族, 私, 周り, トイレ, 家
作業者1	自宅, 家, 自分, 部屋, 職場
作業者2	自分, 私, 周辺, 室内, 家内
作業者3	国, 家庭, 町, 街, 周り
(C) [MASK]軽さから、忘れる軽さへ。	
正解	許せる
ベースライン	忘れる, 壊れる, 思い出す, 嬉しい, 積み重ねる
提案手法	忘れる, 思い出す, 覚える, 思い出す, 気づかない
作業者1	持てる, 運べる, 重ねられる, 運搬できる, 探せる
作業者2	感じる, 受ける, 分かる, 感じ取る, 勘づく
作業者3	思い出す, 残る, 驚く

事例として、事例(C)のような文が挙げられる。この事例で正解単語を予測するには、このキャッチコピーの言及対象がメガネ製品であるという事前知識と、その対象に対する推測やイメージを要する。このような事例にも対応するため、文の言及対象に関する追加情報を入力することも重要である。