

# 意味的プロービングデータセットの構築と言語モデルの評価: イタリア語の倒置を例に

今井咲良<sup>1</sup> Giovanni Pasa<sup>2</sup> 小田博宗<sup>2</sup> 折田奈甫<sup>1</sup> 河原大輔<sup>1</sup>

<sup>1</sup> 早稲田大学理工学術院 <sup>2</sup> 東京大学大学院総合文化研究科

sakura\_imai@toki.waseda.jp, {dkw, orita}@waseda.jp

giov.pasa@gmail.com, hiromuneoda@g.ecc.u-tokyo.ac.jp

## 概要

言語特有の現象を利用した意味的プロービングは、言語モデルにとって挑戦的な検証タスクであり、今後の言語モデルの発展のために必要不可欠である。しかし、英語以外の言語での実施が少ない状況にある。本研究では、多言語への展開を目的として、イタリア語の主語・述語の倒置に着目した意味的プロービングデータセットを作成し、事前学習済み BERT がどの程度の意味識別能力を有するかを観察する。イタリア語の容認性判断タスクを行い、文の持つ統語・意味的性質に対する容認性識別能力の変化を調査した結果、不定代名詞や再帰代名詞など意味解釈と関わりが深い性質をもつ文の容認性識別能力が大きく変動することがわかった。

## 1 はじめに

近年、大規模言語モデルの発展が著しく、様々な自然言語処理タスクにおいて高い精度を達成している。しかし、その高い精度を出すための原理には未だ解明されていない点が多く、いかにしてタスクを解いているのかが注目されている。特に、言語モデルが統語的・意味的にどの程度の言語知識を有しているのか多くの検証が行われている [1, 2, 3]。言語モデルの統語的知識を評価するためのベンチマークに、CoLA [4] や BLiMP [5] などがある。これらは文法的な容認性を二値で判断するタスクであり、英語や日本語をはじめとして様々な言語で構築されている。また、言語モデルの意味に関する知識を評価するベンチマークとしては、文脈中の単語の意味判別を行う WiC [6] などがある。

言語モデルの意味的評価では、言語特有の構文や特徴を用いた検証も行われているが、このような検証は英語での実施に限られる [7, 8]。しかし、英語

以外の言語でも、言語モデルにとって挑戦的な言語固有の構文や現象がある。本研究では、言語モデルの学習や評価に広く用いられる英語以外の言語において言語モデルの意味的判断を検証するため、イタリア語の主語と述語の倒置に焦点を当てる。文中で使用される単語は非倒置文と同じだが、倒置によって語順が変化することで文の意味が変化する。倒置は様々な言語で使用される、自然言語一般に備わる性質である。

本研究では、言語モデルが倒置により生じる意味の変化を識別できるか確かめるため、イタリア語の主語・述語倒置データセットを作成し、意味的プロービングを行う。言語間転移学習によりどの程度の意味的識別能力が転移されるかを検証するため、イタリア語で事前学習されたモデルだけでなく、他の言語で事前学習されたモデルも用いて評価する。また、イタリア語版 CoLA データセット ItaCoLA [9] を用いた容認性判断のタスクを行い、文の持つ統語・意味的性質に対する容認性識別能力の変化を調査する。その結果、イタリア語への言語転移を行った他言語の事前学習モデルは、再帰代名詞や不定代名詞など意味解釈と関わりが深い性質をもつ文の識別能力が変動しやすいことがわかった。

## 2 関連研究

意味的プロービングとは、言語モデルが意味的な識別能力をどれほど有しているかを検証するためのタスクである。英語を対象とした意味的プロービング研究として、Weissweiler ら [7] は、比較相関構文を取り上げ、言語モデルが比較相関構文と似た文との文意の違いを識別できるか検証した。言語モデルは比較相関構文の構造を認識する一方で、その意味的識別精度は低かった。Radford ら [10] は、ゼロショットの GPT-2 では、対話形式の質問応答タスク

など、初歩的な読解能力が必要なタスクでは教師あり学習を行ったベースラインと遜色ない性能を示すが、要約などの意味的な理解が必要なタスクでは不十分であることを確認した。また、言語モデルの意味習得における限界に焦点を当てた研究 [11] では、否定などの自然言語の意味表現の識別能力は、事前学習では獲得できていないことを示した。

このように、意味的な識別能力を必要とするタスクは、近年その発展が著しい言語モデルにとっていまだに挑戦的であり、今後の言語モデル、ひいては自然言語処理の発展のためには必要不可欠な評価である。一方、現時点でデータセットの構築やモデルの評価は英語での実施に限られており、十分に検証が行われているとは言い難い。多言語での実施とさらなる検証が必要である。

### 3 文意判別能力を測る課題の提案

本研究では、新たな意味的プロービング課題の提案にあたり、イタリア語の「主語と述語の倒置」現象に着眼する。主語と述語の倒置は、文中で使用される単語は非倒置文と同じだが、語順によって意味が異なることから、単なる単語の構成性以上の情報が含まれている。このような語順による意味の変化は英語をはじめとした様々な言語に強調などの用法として見られ、自然言語一般に備わる性質である。意味的プロービングで主に使用されてきた英語以外の言語でも、言語モデルがこの種の意味の違いを判別する能力を獲得することができるのか検証するため、本研究ではイタリア語の倒置を意味的プロービング課題として選択する。

#### 3.1 イタリア語の主語と述語の倒置

イタリア語における主語と述語の倒置は、口語的な日常会話で頻繁に用いられ、強調などの効果を持つ。特に、例 (1) のような近過去の時制の文では頻繁に倒置が用いられ、二文とも実質的に同じ意味の文として知覚される。しかし、特定の種類の単語を用いた場合やイディオムとして用いられる表現では、例 (2) のように倒置文と非倒置文において聞き手に受け取られる意味が大きく変化する場合がある。これは、動詞に自動詞用法と他動詞用法の両方がある場合や、他動詞を用いる文において目的語が省略された場合に起こりやすい。本研究では、主語と述語の倒置により解釈されやすい文の意味が強調の範囲を超えて変化しない文ペアを**変化しないセッ**

**表 1** OpenAI gpt-3.5-turbo API で倒置文を生成する指示

Please output pairs of simple Italian sentences followed by their English translations. The order should be: Italian sentence 1, Italian sentence 2, English translation.

Each sentence must consist only of an intransitive verb and a subject.

Each Italian sentence pair consists of two sentences, where in the first the verb comes after the subject, while the second presents an inversion in the order of the verb and subject.

output: Maria ha telefonato., Ha telefonato Maria., Maria called.

トとし ((1) を参照)、それ以上に文の意味が変化する文ペアを**変化するセット**とする ((2) を参照)。

- (1) a. Il pacco è arrivato.  
the package is arrived  
“The package has arrived.”
- b. È arrivato il pacco.  
is arrived the package  
“The package has arrived.”
- (2) a. Il cane mangia.  
the dog eats  
“The dog eats.”
- b. Mangia il cane.  
eats the dog  
“(She/He) Eats the dog.”

#### 3.2 イタリア語主語・述語倒置データセットの作成

本研究では、倒置の有無による当該の意味の変化を言語モデルが識別できるか確かめるため、評価用のデータセットを作成する。倒置文データセットは、文ペアが同じ意味だと解釈される**変化しないセット**、文ペアが違う意味だと解釈される**変化するセット**で構成され、テンプレートおよび大規模言語モデルの API を用いて作成する。

**テンプレートによる作成** 主語、述語、目的語に対して単語を選定し、それらを組み合わせることで文を生成する。倒置による意味の変化は主に動詞に起因するため、述語スロットの動詞候補はイタリア語ネイティブ話者が選定した。主語や目的語は、Marvin ら [12] が容認性判断データの構築に使用したテンプレートから抜粋する。

**API による作成** OpenAI gpt-3.5-turbo モデル<sup>1)</sup>の API を使用し、倒置文の文ペアを生成する。API に与えた指示と生成例を表 1 に示す。

**イタリア語母語話者による確認** テンプレートと API を用いて生成した文ペアは、大学院レベルの言

1) <https://platform.openai.com/docs/models/gpt-3-5>

語学のトレーニングを受けたイタリア語母語話者が全て確認した。基本的な確認項目は、(i) 文ペアに主語と述語の倒置が含まれており、文法的に正しいこと、(ii) 文を単独または特定の文脈下で使用することができることである。それらに加え、**変化しないセット**では、倒置の有無により文ペアで解釈される意味が変化しないこと、**変化するセット**では解釈される意味が変化することを確認する。これらの過程を経て、**変化しないセット** 608 ペア、**変化するセット** 791 ペアを作成した。

## 4 実験

本研究では、言語モデルの意味的判別能力と統語的判別能力を評価するための2種類の実験を行う。意味的判別能力を測る実験では、3.2節で述べたイタリア語の主語・述語倒置データセットを用いる。統語的判別能力を測る実験では、イタリア語版 CoLA データセット ItaCoLA [9] を用いる。各実験は2段階で構成する。はじめにイタリア語データでモデルを追加学習し、次に評価実験を行う。

### 4.1 主語・述語倒置文ペアの意味判別実験

**実験設定** 本実験では、文ペアに対して言語モデルから文脈を捉えたベクトルを取得し、文ペアの類似度を測ることにより意味的プロービングを行う。はじめに、イタリア語の文脈を捉えたベクトルを取得するため、事前学習済み BERT モデルをイタリア語データで追加学習する。この追加学習は、Sentence-BERT [13] の手法を用いたファインチューニングである。その後、3.2節で構築したイタリア語の主語・述語倒置データセットを用いて倒置文ペアの意味判別実験を行う。使用するモデルはイタリア語、スペイン語、英語、ドイツ語の事前学習済み BERT モデルである。詳細を付録 A に示す。

**追加学習** イタリア語の文脈ベクトルを得るための追加学習は、Sentence-BERT [13] のファインチューニングと同様の手法および設定を使用する。使用するデータの設定は、NLI (Natural Language Inference) のみ、STS (Semantic Textual Similarity) のみ、NLI で追加学習を実施した後に STS の追加学習の3パターンを用いる。NLI は、UINAUIL [14] の Textual Entailment セットを使用し、STS は STSb Multi MT [15] のイタリア語セットを使用する。

**主語・述語倒置文ペアの意味判別実験** BERT モデルを追加学習した後、倒置文ペアを用いた意味判

別実験を行う。ペアの各文に対して、追加学習済み Sentence-BERT モデルから全トークンのベクトルを平均した文脈ベクトルを取得する。各文ペアに対して、2つの文脈ベクトルのコサイン類似度を計算する。この類似度を文ペアの意味的な類似度として評価するため、ROC 曲線と AUC を算出する (図 1)。

### 4.2 ItaCoLA の容認性判断実験

**実験設定** 本実験では、事前学習済み GPT-2 モデルをイタリア語 CLM (Causal Language Modeling) タスクで追加学習し、モデルのパープレキシティに基づく容認性判断タスクを行う。使用するモデルはイタリア語、スペイン語、英語、ドイツ語の事前学習済み GPT-2 モデルである。詳細を付録 A に示す。

**追加学習** イタリア語データを用いて GPT-2 モデルを追加学習するために、UINAUIL [14] の Textual Entailment セット (NLI)、および CONcreTEXT [16] (CCT) のテキストを使用する。各データセットを用いて CLM タスクで追加学習する。

**容認性判断実験** 追加学習済みモデルを用いて、ItaCoLA の各文に対してパープレキシティを計算する。このとき、パープレキシティが閾値より小さければ容認可能 (文法的)、大きければ容認不可能 (非文法的) として容認性判断タスクを行い、ROC 曲線と AUC を算出する (図 2)。

## 5 結果と議論

### 5.1 主語・述語倒置文の意味識別実験

図 1 に各言語の事前学習済み BERT モデルに対して、NLI のみ、STS のみで学習した場合、および NLI 学習後に STS で学習した場合において主語・述語倒置データセットを用いて意味的プロービングを行った際の ROC 曲線を示す。

本実験では、イタリア語テキストを用いた追加学習の手法を3種類比較した。Sentence-BERT [13] のファインチューニング設定から、NLI 学習後に STS 学習を行う手法で最も効果的に主語・述語倒置文の意味的判別能力を獲得できるのではないかと推測された。しかし、図 1 から、イタリア語への言語転移によって AUC が改善した英語およびスペイン語モデルでは、STS 学習のみの場合に AUC が向上した。これは、複数のデータによる追加学習は、特定の意味的判別能力の向上には必ずしも最も効果的ではないことを示している。また、追加学習による AUC

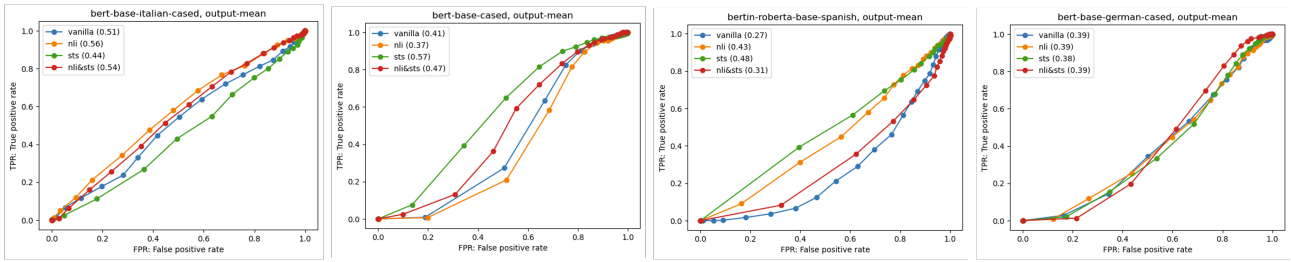


図1 イタリア語の主語・述語倒置文を用いた意味判別実験の結果。左からイタリア語、英語、スペイン語、ドイツ語で事前学習されたBERTモデル。

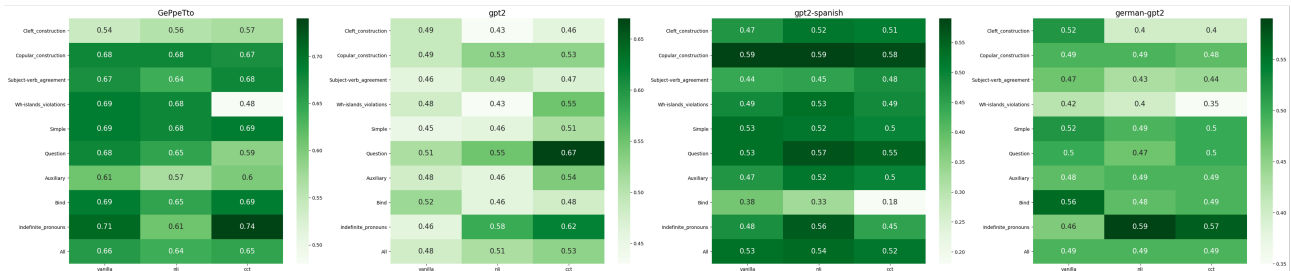


図2 パープレキシティに基づくItaCoLA容認性判断タスクの結果。左からイタリア語、英語、スペイン語、ドイツ語で事前学習されたGPT-2モデル。

の変動は、モデルが事前学習された言語によって異なる。英語、スペイン語モデルでは、イタリア語の追加学習によってAUCが向上しているが、ドイツ語モデルではほぼ変化がない。これは、イタリア語との類似性が関係していると推測される。

個々の文ペアのコサイン類似度を観察する。英語モデルの**変化するセット**では、例(3)のように主語、動詞、目的語の3つの要素から構成され、*raggiungere* “to reach” や *guardare* “to look at” など自動詞・他動詞用法を両方持つ動詞を含む文ペアにおいてコサイン類似度が大きくなりやすい傾向があった。この種の文は、動詞と目的語の連続が主語と動詞の連続より強く意識されるため、倒置により意味が変化しやすい。しかし、言語モデルはこのような意味の変化を捉えることが難しいと考えられる。

- (3) a. Gli insegnanti guardano le ragazze.  
the teachers look the girls  
“The teachers look at the girls.”
- b. Guardano gli insegnanti le ragazze.  
look the teachers the girls  
“The girls look at the teachers.”

## 5.2 ItaCoLAの容認性判断実験

図2に、各言語の事前学習済みGPT-2モデルに対して、追加学習なし(vanilla)、NLI、CCTの設定においてItaCoLA容認性判断を行ったAUCスコアを示す。非イタリア語モデルのうち、全体スコアはスペ

イン語、英語、ドイツ語モデルの順で下がっている。したがって、ItaCoLAを用いた評価においても言語間の類似度の影響が推測され、イタリア語の倒置文を用いた意味判別実験の結果と一貫している。

非イタリア語モデルでは、特に再帰代名詞と不定代名詞の特徴を持つ文で追加学習によるAUCの変動が大きい(図2)。ItaCoLAの不定代名詞の特徴を持つ容認不可能な文は、文法は正しいが語彙の組み合わせが不適切なものが多いことが観察された(付録(5b))。したがって、このAUCの変動は、イタリア語の追加学習を行うことにより、非イタリア語モデルの語彙の意味的な適切さを判別する能力が向上することを示唆している。しかし、ItaCoLAの不定代名詞を含むとされる文に、不定代名詞が含まれず、不定冠詞のみが含まれている例が複数確認されており、これが結果に影響している可能性がある。

## 6 おわりに

本研究では、イタリア語の主語・述語倒置に着目した意味的プロービングデータセットの作成およびプロービング評価を行った。実験では、ItaCoLAを用いたパープレキシティに基づく容認性判断の評価を行い、事前学習済みモデルの意味的判別能力の獲得について検証した。本研究で選択したイタリア語の倒置以外にも、言語モデルに対して挑戦的になり得る構文や現象がある。さらなる言語モデルの検証と多言語への展開は、今後の課題としたい。

## 謝辞

本研究は JSPS 科研費 JP21H04901 の助成を受けて実施した。

## 参考文献

- [1] Lining Zhang, Mengchen Wang, Liben Chen, and Wenxin Zhang. Probing GPT-3’s linguistic knowledge on semantic tasks. In Jasmijn Bastings, Yonatan Belinkov, Yanai Elazar, Dieuwke Hupkes, Naomi Saphra, and Sarah Wiegrefe, editors, **Proceedings of the Fifth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP**, pp. 297–304, Abu Dhabi, United Arab Emirates (Hybrid), December 2022. Association for Computational Linguistics.
- [2] Ethan Wilcox, Roger Levy, Takashi Morita, and Richard Futrell. What do RNN language models learn about filler-gap dependencies? In Tal Linzen, Grzegorz Chrupala, and Afra Alishahi, editors, **Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP**, pp. 211–221, Brussels, Belgium, November 2018. Association for Computational Linguistics.
- [3] Aaron Mueller, Garrett Nicolai, Panayiota Petrou-Zeniou, Natalia Talmina, and Tal Linzen. Cross-linguistic syntactic evaluation of word prediction models. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 5523–5539, Online, July 2020. Association for Computational Linguistics.
- [4] Alex Warstadt, Amanpreet Singh, and Samuel R. Bowman. Neural network acceptability judgments. **Transactions of the Association for Computational Linguistics**, Vol. 7, pp. 625–641, 2019.
- [5] Alex Warstadt, Alicia Parrish, Haokun Liu, Anhad Mohananey, Wei Peng, Sheng-Fu Wang, and Samuel R. Bowman. BLiMP: The benchmark of linguistic minimal pairs for English. **Transactions of the Association for Computational Linguistics**, Vol. 8, pp. 377–392, 2020.
- [6] Mohammad Taher Pilehvar and Jose Camacho-Collados. WiC: the word-in-context dataset for evaluating context-sensitive meaning representations. In Jill Burstein, Christy Doran, and Tamar Solorio, editors, **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. 1267–1273, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [7] Leonie Weissweiler, Valentin Hofmann, Abdullatif Köksal, and Hinrich Schütze. The better your syntax, the better your semantics? probing pretrained language models for the English comparative correlative. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**, pp. 10859–10882, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.
- [8] Lindia Tjuatja, Emmy Liu, Lori Levin, and Graham Neubig. Syntax and semantics meet in the “middle”: Probing the syntax-semantics interface of LMs through agentivity. In Alexis Palmer and Jose Camacho-collados, editors, **Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (\*SEM 2023)**, pp. 149–164, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [9] Daniela Trotta, Raffaele Guarasci, Elisa Leonardelli, and Sara Tonelli. Monolingual and cross-lingual acceptability judgments with the Italian CoLA corpus. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, **Findings of the Association for Computational Linguistics: EMNLP 2021**, pp. 2929–2940, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [10] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.
- [11] Nora Kassner and Hinrich Schütze. Negated and misprimed probes for pretrained language models: Birds can talk, but cannot fly. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 7811–7818, Online, July 2020. Association for Computational Linguistics.
- [12] Rebecca Marvin and Tal Linzen. Targeted syntactic evaluation of language models. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun’ichi Tsujii, editors, **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**, pp. 1192–1202, Brussels, Belgium, October–November 2018. Association for Computational Linguistics.
- [13] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 3982–3992, Hong Kong, China, November 2019. Association for Computational Linguistics.
- [14] Valerio Basile, Livio Bioglio, Alessio Bosca, Cristina Bosco, and Viviana Patti. UINAUIL: A unified benchmark for Italian natural language understanding. In Danushka Bollegala, Ruihong Huang, and Alan Ritter, editors, **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)**, pp. 348–356, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [15] Philip May. Machine translated multilingual sts benchmark dataset. 2021.
- [16] Lorenzo Gregori, Maria Montefinese, Daniele Paolo Radicioni, Andrea Amelio Ravelli, and Rossella Varvara. Concretex @ evalita2020: The concreteness in context task. **EVALITA Evaluation of NLP and Speech Tools for Italian - December 17th, 2020**, 2020.

## A 学習および評価に使用したモデルの詳細

イタリア語の主語・述語倒置の意味判別実験に使用するモデルはイタリア語<sup>2)</sup>、スペイン語<sup>3)</sup>、英語<sup>4)</sup>、ドイツ語<sup>5)</sup>の事前学習済み BERT モデルである。また、ItaCoLA の容認性判断実験に使用するモデルはイタリア語<sup>6)</sup>、スペイン語<sup>7)</sup>、英語<sup>8)</sup>、ドイツ語<sup>9)</sup>の事前学習済み GPT-2 モデルである。

## B 議論

本節では、5 節で述べた結果に関する議論を補足する。

### B.1 主語・述語倒置文の意味識別実験

個々の文ペアのコサイン類似度を観察すると、モデルが事前学習された言語に関わらず、単語数の少ない現在時制の文ペアでコサイン類似度が低くなりやすく、単語数の多い文ペアではコサイン類似度が高くなりやすい傾向があった。変化しないセットにおける、短い文 (4a) と長い文 (4b) の例を示す。単語数の多い長い文では場所を表す句等が追加されている場合が多い。変化しないセットでは動詞の性質によって意味の変化が生じる場合が多いため、文長に対して明らかに同じ意味を表す部分が多いことからコサイン類似度が高まりやすいことが推測される。

- (4) a. I manager ridono.  
“The managers laugh.”
- b. Il chirurgo soggiorna a casa.  
“The surgeon stays at home.”

### B.2 ItaCoLA の容認性判断実験

5.2 節で述べたように、ItaCoLA の不定代名詞の特徴を持つ文では、不適切な語彙の組み合わせにより容認不可能となる例が観察された。(5a) に語彙の組み合わせが適切な例、(5b) に語彙の組み合わせが不適切な例を示す。

表 2 イタリア語モデルとスペイン語モデルのイタリア語テキストのトークナイズの違い

---

テキスト: Se stesso inganna Lorenzo. (Lorenzo deceives himself.)
イタリア語モデル: ['Se', 'stesso', 'inganna', 'Lorenzo', '.']
スペイン語モデル: ['Se', 's', 'tes', 'so', 'in', 'gan', 'na', 'Lorenzo', '.']

---

- (5) a. Sofia si è pigliata una forte cotta per Riccardo.  
“Sofia had a strong crush on Richard.”
- b. \*La zanzara si è pigliata una forte cotta per Riccardo.  
“The mosquito had a strong crush on Richard.”

個々のモデルに対して、ItaCoLA の文の特徴とパープレキシティの値を観察する。イタリア語の再帰代名詞と同様の用法があるスペイン語のモデルは、vanilla では、非文である再帰代名詞 *se stesso* “oneself” と自動詞が共存する文に対して高いパープレキシティを付与しているため、再帰代名詞を他動詞とのみ使用できることを理解していると推測される。しかし、CCT で追加学習したモデルでは、再帰代名詞と自動詞の組み合わせに対してパープレキシティを比較的lowめに出力している。これは、イタリア語モデルとスペイン語モデルのトークナイズの違いに起因しているのではないかと考えられる。表 2 にモデルによるイタリア語テキストのトークナイズの結果を示す。イタリア語モデルでは、*se stesso* を正しく単語区切りしているのに対し、スペイン語モデルでは *stesso* を文字区切りにしている。したがって、トークナイズの違いにより追加学習でイタリア語テキストの学習が逆効果になっていることを示している。

2) <https://huggingface.co/dbmdz/bert-base-italian-cased>

3) <https://huggingface.co/bertin-project/bertin-roberta-base-spanish>

4) <https://huggingface.co/bert-base-cased>

5) <https://huggingface.co/bert-base-german-cased>

6) <https://huggingface.co/LorenzoDeMattei/GePpeTto>

7) <https://huggingface.co/DeepESP/gpt2-spanish>

8) <https://huggingface.co/gpt2>

9) <https://huggingface.co/dbmdz/german-gpt2>