► Evaluating Synonym and Antonym Acquisition from a Portuguese Masked Language Model

Hugo Gonçalo Oliveira,
CISUC, Department of Informatics Engineering
University of Coimbra, Coimbra, Portugal
hroliv@dei.uc.pt

As for many other tasks in the domain of Natural Language Processing, transformer-based models have been explored in the acquisition of semantic relations, and are useful for the automatic creation or enrichment of knowledge bases [1], e.g., represented in RDF. Towards this, they can be used for completing lexico-syntactic patterns, in a shortcut to earlier methods of relation acquisition from corpora [12]. When focused on lexico-semantic relations, they can be useful for enriching lexical knowledge bases (LKBs) like WordNet [4].

For Portuguese, BERT [3] has been used for detecting hyponymy pairs [14] and discovering arguments of lexico-semantic relations [9]. Here, we adopt the previous method, but focus on synonymy and antonymy, both symmetric relations that may pose different challenges. In what can be seen as a zero-shot learning approach, we prompt BERTimbau [15] with masked lexical patterns that transmit the target relations, and rely on the model predictions for the masks. For evaluation, we propose two available tests of Portuguese lexico-semantic knowledge: B2SG [16] and TALES [10].

B2SG is similar to the WordNet-Based Synonymy Test [5], but based on the Portuguese part of BabelNet [13] and partially evaluated by humans. It contains Portuguese words (source) followed by four candidates, out of which only one is related, and is organised in three relations: synonymy, antonymy, and hypernymy between nouns and verbs, respectively. An example for noun-synonymy is: cataclismo desastre_noun talha_noun obesidade_noun alusão_noun (cataclysm disaster obesity allusion).

TALES [10] was created from the contents of ten lexical resources for assessing lexico-semantic analogies in Portuguese, and its format is similar to BATS [7]. For each relation, TALES includes 50 entries with two columns: a word (source) and a list of related words (targets). An example for antonym-of is: `novo velho/idoso/entradote` (young old/aged/oldish).

Given the source words, several handcrafted lexical patterns indicating synonymy and antonymy were tested with two versions of BERTimbau (base and large) for selecting the related words in B2SG, and predicting the target words in TALES. For BSG, the process is simplified with FitBERT,[1] a tool that, given a masked sentence and a

[1] https://github.com/Qordobacode/fitbert

list of options, selects the most suitable option for the mask, based on pre-softmax logit scores [8] computed on BERT. Performance in both tests can be measured by accuracy. In BSG, we also look at the average ranking of the correct answer (1–4), and in TALES at the presence of correct words in the top-10 predictions, i.e., Accuracy@10. The tested patterns included those from the relation-validation service VARRA [6], which were among the best performing.

Table 1 displays the best performing patterns for each relation and test. From the accuracy in B2SG, we confirm that, despite being far from perfect, answers are also far from random, which would yield 0.25. A preliminary analysis suggests that BERT-large is preferable for the majority of relations, and that the procedure may suit antonymy better. The restricted number of antonyms of a word may contribute to this.

When compared to the performance for other relations, the best performance in BSG is for noun-antonymy, even more accurate than noun-hypernymy (0.64), while verb-synonymy is less accurate, but still higher than verb-hypernymy (0.52). In TALES, comparison is more difficult, because accuracies are typically lower and the test covers different types of hypernymy and hyponymy.

Due to its connection with semantic similarity, an initial hypothesis was that synonymy is better captured by similarity-based methods than with a single lexical pattern. So, we tested three different approaches for answering B2SG by selecting the candidate that maximises similarity with the source, based on embeddings from: the `CLS` token of the pretrained BERTimbau; BERTimbau fine-tuned for Natural Language Inference[2]; and GloVe pretrained for Portuguese [11].[3] Accuracy was worse with the pretrained model, except for noun-synonymy (0.67); with the NLI model, there were minor improvements for all but verb-antonymy; while GloVe clearly outperformed the pattern-based approach, with accuracies between 0.7 (noun-synonymy) and 0.83 (noun-antonymy). This supports the initial hypothesis, not only for synonymy, but also antonymy.

In the future, we will deepen this analysis and use the same datasets for evaluating the acquisition of lexico-semantic relations from neural language models. A similar analysis, based on the completion of patterns, may also be performed in generative models like GPT-3 [2] or similar ones that recently became available [17].

---

2 `ricardo-filho/bert-portuguese-cased-nli-assin-assin-2` in HuggingFace.

3 <1% of the words in the test were not covered by the vocabulary of this model.

Table 1: Best performing patterns for each relation in each test.

B2SG						
Relation	**PoS**	**Pattern**	**BERT-base**		**BERT-large**	
			Acc	**Rank**	**Acc**	**Rank**
Synonym-of	N	`é o mesmo que [MASK]`	0.57	1.71	0.64	1.58
Synonym-of	V	`, isto é, [MASK]`	0.50	1.80	0.56	1.67
Antonym-of	N	`nem [MASK], nem`	0.76	1.64	0.77	1.36
Antonym-of	V	`nem [MASK], nem`	0.63	1.64	0.61	1.61
TALES						
Relation	**PoS**	**Pattern**	**BERT-base**		**BERT-large**	
			Acc	**Acc@10**	**Acc**	**Acc@10**
Synonym-of	N	`é sinónimo de [MASK]`	0.28	0.64	0.20	0.70
Synonym-of	V	`é o mesmo que [MASK]`	0.12	0.80	0.34	0.90
Synonym-of	ADJ	`estar é o mesmo que estar [MASK].`	0.06	0.46	0.24	0.54
Antonym-of	ADJ	`ser [MASK] é o contrário de ser`	0.26	0.40	0.38	0.48

Acknowledgements: This work was partially supported by: the COST Action CA18209 Nexus Linguarum (European network for Web-centred linguistic data science); national funds through the FCT – Foundation for Science and Technology, I.P., within the scope of the project CISUC – UID/CEC/00326/2020 and by the European Social Fund, through the Regional Operational Program Centro 2020.

References

1. AlKhamissi, B., Li, M., Celikyilmaz, A., Diab, M., Ghazvininejad, M.: A review on language models as knowledge bases. https://arxiv.org/abs/2204.06031 (2022)

2. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D.: Language models are few-shot learners. In: Advances in Neural Information Processing Systems. vol. 33, pp. 1877–1901. Curran Associates, Inc. (2020)

3. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proc 2019 Conf of North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 4171–4186. ACL (2019)

4. Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database (Language, Speech, and Communication). The MIT Press (1998)

5. Freitag, D., Blume, M., Byrnes, J., Chow, E., Kapadia, S., Rohwer, R., Wang, Z.: New experiments in distributional representations of synonymy. In: Procs 9th Conference on Computational Natural Language Learning (CoNLL-2005). pp. 25–32. ACL, Ann Arbor, Michigan (2005)

6. Freitas, C., Santos, D., Gonçalo Oliveira, H., Quental, V.: VARRA: Validação, Avaliação e Revisão de Relações semânticas no AC/DC. In: Pesquisas e perspetivas em linguistica de corpus (Livro do IX Encontro de Linguística de Corpus, 2010), pp. 199–232. ELC 2010, Mercado de Letras, Rio Grande do Sul, Brasil (2015)

7. Gladkova, A., Drozd, A., Matsuoka, S.: Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn't. In: Procs of NAACL 2016 Student Research Workshop. pp. 8–15. ACL (2016)

8. Goldberg, Y.: Assessing BERT's syntactic abilities. https://arxiv.org/abs/1901.05287 (2019)

9. Gonçalo Oliveira, H.: Drilling lexico-semantic knowledge in Portuguese from BERT. In: Computational Processing of the Portuguese Language – 14th International Conf, PROPOR 2022. LNCS, vol. 12037, pp. 387—397. Springer (2022)

10. Gonçalo Oliveira, H., Sousa, T., Alves, A.: TALES: Test set of Portuguese lexical-semantic relations for assessing word embeddings. In: Procs of ECAI 2020 Workshop on Hybrid Intelligence for Natural Language Processing Tasks (HI4NLP 2020). CEUR Workshop Proceedings, vol. 2693, pp. 41–47. CEUR-WS.org (2020)

11. Hartmann, N.S., Fonseca, E.R., Shulby, C.D., Treviso, M.V., Rodrigues, J.S., Aluísio, S.M.: Portuguese word embeddings: Evaluating on word analogies and natural language tasks. In: Proc of 11th Brazilian Symposium in Information and Human Language Technology (STIL 2017) (2017)

12. Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: Proc 14th Conference on Computational Linguistics. pp. 539–545. COLING 92, ACL (1992)

13. Navigli, R., Ponzetto, S.P.: BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artificial Intelligence 193, 217–250 (2012)

14. Paes, G.E.: Detecção de Hiperônimos com BERT e Padrões de Hearst. Master's thesis, Universidade Federal de Mato Grosso do Sul (2021)

15. Souza, F., Nogueira, R., Lotufo, R.: BERTimbau: Pretrained BERT models for

Brazilian Portuguese. In: Proc of Brazilian Conf on Intelligent Systems (BRACIS 2020). LNCS, vol. 12319, pp. 403–417. Springer (2020)

16. Wilkens, R., Zilio, L., Ferreira, E., Villavicencio, A.: B2SG: a TOEFL-like task for Portuguese. In: Procs 10th International Conference on Language Resources and Evaluation (LREC 2016). ELRA (may 2016)

17. Zhang, S., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., Dewan, C., Diab, M., Li, X., Lin, X.V., Mihaylov, T., Ott, M., Shleifer, S., Shuster, K., Simig, D., Koura, P.S., Sridhar, A., Wang, T., Zettlemoyer, L.: OPT: Open pre-trained transformer language models. https://arxiv.org/abs/2205.01068 (2022)