

# Integrating Semantic Information into Sketchy Reading Module of Retro-Reader for Vietnamese Machine Reading Comprehension

Hang Thi-Thu Le<sup>1,2</sup>, Viet-Duc Ho<sup>1,2</sup>, Duc-Vu Nguyen<sup>\*,1,2</sup>, Ngan Luu-Thuy Nguyen<sup>1,2</sup>

<sup>1</sup>University of Information Technology, Ho Chi Minh City, Vietnam

<sup>2</sup>Vietnam National University, Ho Chi Minh City, Vietnam

{18520274, 18520610}@gm.uit.edu.vn

{vund, ngannlt}@uit.edu.vn

**Abstract**—Machine Reading Comprehension has become one of the most advanced and popular research topics in the fields of Natural Language Processing in recent years. The classification of answerability questions is a relatively significant sub-task in machine reading comprehension; however, there haven't been many studies. Retro-Reader is one of the studies that has solved this problem effectively. However, the encoders of most traditional machine reading comprehension models in general and Retro-Reader, in particular, have not been able to exploit the contextual semantic information of the context completely. Inspired by SemBERT, we use semantic role labels from the Semantic Role Labeling (SRL) task to add semantics to pre-trained language models such as mBERT, XLM-R, PhoBERT. This experiment was conducted to compare the influence of semantics on the classification of answerability for the Vietnamese machine reading comprehension. Additionally, we hope this experiment will enhance the encoder for the Retro-Reader model's Sketchy Reading Module. The improved Retro-Reader model's encoder with semantics was first applied to the Vietnamese Machine Reading Comprehension task and obtained positive results.

**Index Terms**—Machine Reading Comprehension, Semantic Role Labeling, SemBERT, Natural Language Processing, Natural Language Understanding

## I. INTRODUCTION

Machine reading comprehension (MRC) is one of the challenging tasks of natural language processing; This task requires the model to answer questions based on specific passages [1], [2]. In order to give accurate answers to the questions in the MRC task, the model needs to be able to read and understand the model language at a high level. In recent years, the MRC's model must not only deal well with giving correct answers [3], [4] to questions but also distinguish unanswerable questions to avoid giving inappropriate answers [5].

Determining the ability to answer in machine reading comprehension tasks is essential; earlier studies have partly solved this problem in English [6], [7]. Typically, the Retro-Reader improvement model of Zhang et al. [8] has obtained state-of-the-art results compared to previous studies.

According to [8], so far, standard reading systems consist of two modules: 1) an encoder with a robust language model and

2) a decoder cleverly designed to match the task characteristics of the MRC. Retrospective Reader (Retro-Reader) has designed a reading comprehension model that enhances the classification problem by integrating two phases of reading and verification strategies. However, this model has been too focused on the decoder and not enough on the encoder aspect.

Pre-trained language models (PrLMs) such as ELMo [9], GPT [10], BERT [11], RoBERTa [12], ALBERT [13], which have succeeded in various natural language processing tasks, are well-known and play the role of a powerful encoder. However, existing language models only focus on exploiting the plain contextual features, rarely considering explicit contextual semantic clues. That may be one of the shortcomings of current language representation models.

By integrating the plain context representation on BERT and explicit semantics from semantic role labels for deeper meaning representation, Zhang et al. [14] obtained state-of-the-art results compared to regular BERT. This has demonstrated that providing explicit semantics to pre-trained language models is necessary.

Many studies in English utilize semantic information from semantic role labels to improve Natural Language Understanding (NLU) tasks. However, there is very little research on SRL tasks in Vietnamese, and those that exist mainly concentrate on developing semantic labeling systems [15]–[17] without using the semantics of semantic role labels for other tasks. That is the motivation for us to do this study.

Inspired by SemBERT [14] and motivated by the aim to improve the encoder for the Retro-Reader model, we propose to implement a Retro-Reader model that includes semantic information to compare the influence of semantics on the classification of answerability for Vietnamese MRC. So, we integrate semantic information into the encoder of the classifier module in Retro-Reader.

The primary contributions of our current research:

- Comparing the influence of semantics from semantic role labeling task on classifying answerability of questions for Vietnamese machine reading comprehension.
- Improved encoder for Sketchy Reading Module of Retro-Reader model.

\*Corresponding author

- Demonstrate the significance of semantic role labels in Vietnamese.

## II. RELATED WORKS

### A. Machine Reading Comprehension

The origins of the MRC extend back to the 1970s when researchers started to recognize the significance of text understanding in artificial intelligence. The first reading comprehension system proposed by W Lehnert in 1977, QUALM [18], established the foundation for developing later machine reading comprehension research.

Machine reading comprehension was clearly described and handled as a supervised learning task by 2013: given a passage and a relevant question, the model should give the answer in the required format. The MRC task has long been in development in English, with the early introduction of datasets [5], [19]. In contrast, the MRC problem in Vietnamese has only begun to receive attention, with several datasets such as UIT-ViQuAD1.0 [20], ViMMRC [21], UIT-ViNewsQA [22]. Along with that, many studies on Vietnamese MRC systems have been developed [23], [24]. One problem encountered by the above MRC systems is that these MRC systems are designed with the assumption that all questions can be answered but this is not always the case in actual life. Consequently, the researchers added unanswerable questions to the MRC datasets later on [5], [25], [26]. This is a significant challenge for later MRC models; a successful MRC model must be able to properly handle two aspects: providing accurate answers to answerable questions and identifying the answerability of the question.

Distinguishing unanswerable questions is a relatively important task in Automatic Reading Comprehension, but previous studies have not paid attention to this issue or only used simple methods. Most of the question's answerability classification tasks are trained with answer span prediction. Liu et al. [27] added a [CLS] token to the context, and used an additional layer of simple classifiers to their MRC model, Back et al. [7] relied on the attention mechanism to identify the words in the question that made them impossible to answer, thereby calculating a score for the questions and determining answerability based on that score.

Unlike previous studies, the Retro-Reader [8] model takes an entirely new approach with two strategic phases: read and verify. It obtained state-of-the-art results on the MRC in English. In this study, we will enhance the Retro-Reader model and implement it for the Vietnamese MRC.

### B. Explicit Contextual Semantics

The semantic role labeling task aims to identify the predicate-argument structure of a sentence by analyzing the shallow semantics of natural language documents. It assists us in finding the answers to questions like "Who?" "Did what?" "For whom?" "With what?" "Where?" and "When?" [28], [29]. Coincidentally, those are the same issues that MRC tasks in general or QA in particular have to deal with; it is entirely fair to apply semantic role labels for the MRC task.

In 2004, Narayanan and Harabagiu [30] proposed employing semantic roles in question answering (QA) systems; this was one of the earliest research on using semantic role labels as a side task in QA. Most subsequent studies [31], [32] that used semantic roles as a primary method in the QA process yielded numerous positive outcomes. As a result of the dominance of pre-trained language models [9]–[12], semantic information from semantic role labels is no longer valued or utilized. However, semantic labels are helpful in many other NLP applications [33], [34].

In 2020, Zhang et al. [14] proposed incorporating explicit contextual semantics from the pre-trained semantic role labeling task and introducing a more effective language representation model called SemBERT. It achieves state-of-the-art and significantly improves results on ten reading comprehension and linguistic inference tasks in English.

There has not been any research on Vietnamese using semantic information from semantic labels to enhance MRC/QA tasks or NLU tasks. To compare the impact of semantics on the classification of answerability for the Vietnamese MRC task, we will use SemBERT's method to create variations of SemBERT from several pre-trained language models.

## III. METHODOLOGY

### A. Retro-Reader Model

The retrospective reader (Retro-Reader) [8] has two parallel modules that can be used to execute a two-stage reading process: a *sketchy reading module* and an *intensive reading module*. Then, combine the answerability confidence score (intensive reading module) with the judgment score (sketchy reading module) to yield the final answer called *rear verification*. In both modules, [8] apply BERT-based PrLMs as Encoder to acquire contextual representations of input tokens. An overview of the Retro-Reader model is shown in figure 1.

**Sketchy reading module** can be seen as a classifier to verify the answerability of the question; the final hidden state is passed through a fully connected layer to get classification logits composed of answerable and unanswerable elements. The output score differs between the no-answer score and the answer score.

**Intensive reading module** with the aims of verifying answerability, predicting the answer span, and giving the final predicted answer. Use the final layer hidden state to predict each token's start and end probabilities, thereby giving the predicted answer span and no-answer score. In addition, there is an internal front verifier such that the intensive reader can also identify unanswerable questions.

**Rear verification** combines the output of the two modules above, which is an aggregated verification for the final answer.

### B. Semantic-aware BERT

Semantic-aware BERT (SemBERT) [14] model has three main components, is represented in Figure 3: 1) a semantic role labeler for annotating input sentences with various semantic role labels; 2) A sequence encoder that uses a pre-trained language representation model to produce a contextual representation of

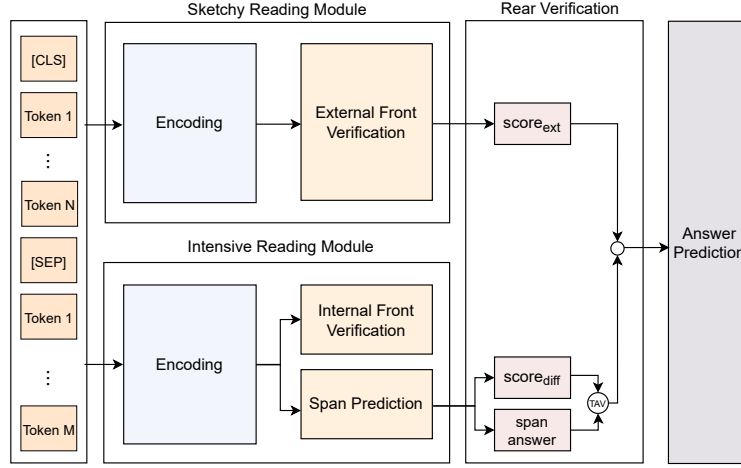


Figure 1. Overview of the Retrospective Reader model [8]

the input raw text and mapped semantic labels in parallel to create a semantic representation; 3) A semantic integration component that combines the text representation with the explicit contextual semantic embedding to produce an integrated representation for downstream tasks.

**Semantic Role Labeling** is the process of assigning semantic roles to input sentences with varying semantic sequences using a pre-trained semantic labeler. There will be many predicate-argument structures for a particular sentence, such as the one in the figure 3.

**Encoding** is a task of learning the contextual representation through the BERT model and semantics representation through the input’s semantic role labels, including two parallel processes.

- **Contextual Embedding:** The BERT model learns the contextual representation of the input text at the subword level, then groups the subword representation word by word and uses a convolutional neural network (CNN) with max pooling to obtain the contextual representation at the word level.
- **Semantic Embedding:** These label sequences obtained from the semantic role labeler are mapped through the lookup table to form vectors and fed to the BiGRU layer to obtain a representation in the latent space, then concatenate and feed them into a fully connected class to get a semantic representation.

**Integration:** Contextual representation at word level and semantic representation based on semantic role labels are concatenated together, forming a common representation of downstream tasks.

#### IV. EXPERIMENTS AND RESULTS

##### A. Dataset

1) **SRL Task:** We use the corpus of the LORELEI Vietnamese Language Pack [35] for the SRL problem. This data is collected from discussion forums, message boards, documents, social networks, and online data, including monolingual and

bilingual text, vocabulary, and annotations. The dataset includes 1760 sentences that have been labeled with Simple Semantic Annotation [36].

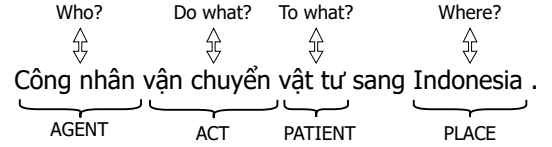


Figure 2. Example of Vietnamese SRL with SSA annotation

##### B. Semantic-aware BERT

2) **MRC Task:** UIT-ViQuAD stands for Vietnamese Question Answering Dataset, which was created specifically for the task of Vietnamese machine reading comprehension based on passages extracted from Vietnamese Wikipedia articles. Initially, UIT-ViQuAD1.0 [20] consisted of only answerable questions. To increase machine learning, the UIT-ViQuAD2.0 [26] combined 23K questions in UIT-ViQuAD1.0 with more than 12K unanswered questions.

We separated the dataset for the MRC task into a training set, a development set, and a test set; the number of each dataset is detailed in the table I.

Table I  
OVERVIEW STATISTICS OF THE MRC DATASET.

	Train	Dev	Test	All
Number of articles	111	14	13	138
Number of passages	3,038	637	426	4,101
Number of questions	21,234	3,100	4,123	28,457
Number of unanswerable questions	6,890	1,359	968	9,217

##### C. Evaluation Metrics

To evaluate the performance of the model in the Vietnamese Machine Reading Comprehension task, we use two measures.

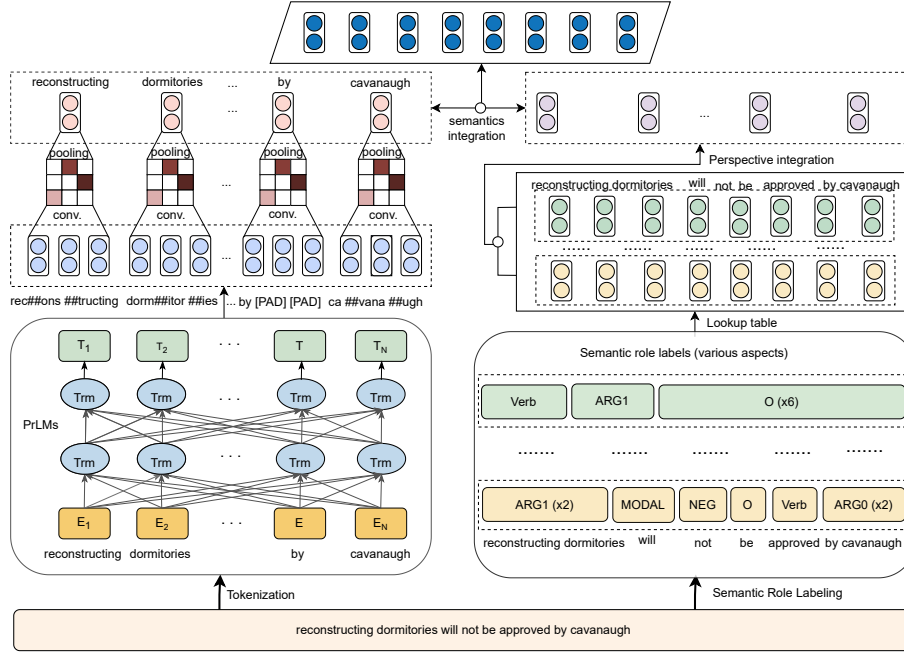


Figure 3. Overview of the Semantic-aware BERT model [14]

**Exact Match (EM)** is the number of precise answers, giving a score of 1 when the prediction and the true answer are the same and 0 otherwise. When evaluating against a negative question, if the system predicts any textual span as an answer, it automatically obtains a zero score for that question.

**F1-score** estimated over the individual tokens in the predicted answer against those in the gold standard answers is based on the number of matched tokens between the expected and gold standard answers.

$$Precision = \frac{N_{matched}}{N_{predicted}}$$

$$Recall = \frac{N_{matched}}{N_{gold\_standard}}$$

$$F1score = \frac{2 * Precision * Recall}{Precision + Recall}$$

where  $N_{matched}$  is the number of matched tokens.  $N_{predicted}$  and  $N_{gold\_standard}$  is the number of the predicted answer tokens and the gold standard answer tokens respectively.

#### D. Fine-tuning SRL model

A sentence may have more than one predicate-argument for the role labeling task. Therefore, we trained ten separate semantic role labeling models using the k-fold cross validation approach with  $k = 10$  to extract various aspects of semantic information from a sentence. For each fold, we fine-tuned the XLM-RoBERTa model on 40 epochs and used AdamW [37] for optimization with a learning rate  $1e - 5$ , weight decay 0.1, and batch size 4.

#### E. Retro Reader

We used monolingual and multilingual pre-training language models, such as mBERT, XLM-RoBERTa, and PhoBERT in the base version.

- For the sketchy reading module, we use the following hyper-parameters learning rate  $5e - 6$  with the AdamW optimizer and batch size 64, which has a cumulative gradient of 8.
- For the intensive reading module, we optimally use AdamW with a learning rate of  $2e - 5$  and batch size 64. The maximum length of input tokens `max_seq_length` is the maximum length of the language representation model, and the `max_query_length` query length is 64.

Finally, to integrate the two above modules, we also try to change the ratio parameter between the two models above; the thresholds are selected based on the dev set.

Additionally, we use SemBERT as a language representation model and incorporate it and its variants into Retro-sketchy Reader's reading module, fine-tuning with SemBERT's final output representation. We employ the Retro-Reader model's hyper-parameters.

### V. RESULTS AND DISCUSSION

Table II shows the performance evaluation of the models on the UIT-ViQuADv2.0 dataset. Accordingly, we can readily see a favorable difference between utilizing semantics and not using semantics for the Retro-Reader model. Using semantics increased the overall model evaluation score from 0.5 - 1% on both F1\_score and EM measures. This improvement is statistically significant at  $p < 0.01$  using paired t-test.

The results of the Sketchy Reading Modules of each model in the answerability question classification task are shown in Table



Table II

TABLE OF RESULTS OF MACHINE READING COMPREHENSION MODELS. THE SYMBOL  $\ddagger$  DENOTES THAT THE IMPROVEMENT IS STATISTICALLY SIGNIFICANT AT  $P < 0.01$  COMPARED WITH RETRO-READER'S RESULTS USING PAIRED T-TEST.

Model		EM (%)	F1-score (%)
MRC Baseline	mBERT	42.130	55.000
	XLM-R	45.234	57.872
	PhoBERT	46.859	60.191
Retro-Reader	mBERT	44.118	55.858
	XLM-R	48.072	60.521
	PhoBERT	48.363	61.262
Retro-Reader + Semantic (CLS)	mBERT	44.967 $\ddagger$	56.326 $\ddagger$
	XLM-R	48.485 $\ddagger$	61.065 $\ddagger$
	PhoBERT	<b>49.032<math>\ddagger</math></b>	<b>61.896<math>\ddagger</math></b>

III. As we can see, applying semantic from semantic role labels to the three language models, mBERT, XLM-RoBERTa, and PhoBERT, yielded an increase of 4.341%, 1.9%, and 1.414%, respectively, compared to the conventional contextual model in the task of identifying unanswerable questions. This helps to improve the total score for the Retro-Reader model on the Machine Reading Comprehension.

Table III

THE RESULTS OF THE SKETCHY READING MODULE'S UNANSWERABLE QUESTION VERIFICATION TASK.

	Accuracy (%)	F1-score (%)
BERT	71.258	48.275
BERT + Semantic	73.490	55.518
XLMR	76.110	62.188
XLMR + Semantic	76.985	66.695
PhoBERT	73.805	53.333
PhoBERT + Semantic	75.036	56.954

According to the results for the SRL task in table IV, we can see that the F1-score of the models is poor, which can be explained by the amount of SRL data in the LORELEI dataset we used being relatively limited and not enough to train a good labeling model. This can influence the labeling of input sentences, and inaccurate semantic labels can damage the extraction of semantic information; thus, we anticipate these improvements could be significantly better if we trained a more accurate SRL model.

Overall, the Retro-Reader model improvement ranges from 0.5% to 1%, which is insignificant for the Vietnamese MRC task, but our experiment has partly shown the potential of semantic extraction from semantic role labeling in Vietnamese. The positive results of this experiment will serve as the basis for future research into the usage of SRL for other NLU tasks in Vietnamese.

## VI. CONCLUSION

### A. Summary

In this study, we mainly compare the effect of semantics on the classification of answerability questions for the Vietnamese

Table IV

THE RESULTS OF THE SRL PROBLEM AFTER FINE-TUNING XLM-R, ON THE LORELEI DATASET.

	Precision (%)	Recall (%)	F1-score (%)
Fold 1	35.261	32.568	33.871
Fold 2	34.700	30.375	32.393
Fold 3	38.776	32.095	35.120
Fold 4	41.958	29.557	34.682
Fold 5	38.821	31.728	34.917
Fold 6	40.144	34.901	34.900
Fold 7	36.719	32.192	34.307
Fold 8	39.535	24.286	36.724
Fold 9	39.241	32.518	35.564
Fold 10	39.448	34.014	36.530
Average	38.460	31.423	34.901

MRC problem by integrating semantic information into the encoder of the sketchy reading module in Retro-Reader. The results show that the base version mBERT, XLM-R, and PhoBERT models with semantic information perform better than conventional contextual representation models. We have described the implementation approach and experimental results for the MRC problem. In particular, the results improved by 1.4 - 4.3% for the Retro-Reader model's Sketchy Reading Module in the sub-task of classifying the question's answerability, helping the model's overall results improve by 0.5 - 1%. This is one of the earliest research utilizing semantic information from semantic labels for language representation in Vietnamese. As a result of the acquired experimental results, it is possible to realize the potential of semantic extraction from SRL for various NLU tasks in Vietnamese.

However, during the implementation of this experiment, we encountered some difficulties in training pre-trained SRL models because of the limitation of the corpus size. This has affected the annotation of semantic role labels for the input sentences. And based on the encouraging results of this study, we want to partially encourage further research on the SRL task for Vietnamese, as there are currently few research works and corpora for this task in Vietnamese.

### B. Future Work

We identify several directions for further research and development based on the results and study limitations, including:

- Due to time and resource limitations, we have only tested the pre-trained language model's base version and a small number of variants. In the future, we will try more semantic-aware BERT model variants with a larger version.
- In this study, we have only focused on improving the answerability classification task, but in the future, we will be testing semantic integration into Retro-Reader's Intensive Reading Module.
- We believe that if there is enough data to train an SRL model, the input sentence labeling will be enhanced, allowing the semantic integration model to perform better. As a result, in the future, we will create a semantic

role labeling dataset to train an SRL model that is more accurate than the existing model.

#### ACKNOWLEDGEMENT

This research is funded by University of Information Technology-Vietnam National University HoChiMinh City under grant number D1-2022-47.

#### REFERENCES

- [1] K. M. Hermann, T. Kocisky, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom, "Teaching machines to read and comprehend," *Advances in neural information processing systems*, vol. 28, 2015.
- [2] M. Joshi, E. Choi, D. S. Weld, and L. Zettlemoyer, "Triviaqa: A large scale distantly supervised challenge dataset for reading comprehension," *arXiv preprint arXiv:1705.03551*, 2017.
- [3] M. Lewis and A. Fan, "Generative question answering: Learning to answer the whole question," in *International Conference on Learning Representations*, 2018.
- [4] J. Lee, S. Yun, H. Kim, M. Ko, and J. Kang, "Ranking paragraphs for improving answer recall in open-domain question answering," *arXiv preprint arXiv:1810.00494*, 2018.
- [5] P. Rajpurkar, R. Jia, and P. Liang, "Know what you don't know: Unanswerable questions for squad," *arXiv preprint arXiv:1806.03822*, 2018.
- [6] M. Hu, F. Wei, Y. Peng, Z. Huang, N. Yang, and D. Li, "Read+verify: Machine reading comprehension with unanswerable questions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 6529–6537.
- [7] S. Back, S. C. Chinthakindi, A. Kedia, H. Lee, and J. Choo, "Neurquri: Neural question requirement inspector for answerability prediction in machine reading comprehension," in *International Conference on Learning Representations*, 2019.
- [8] Z. Zhang, J. Yang, and H. Zhao, "Retrospective reader for machine reading comprehension," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 16, 2021, pp. 14 506–14 514.
- [9] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 2227–2237. [Online]. Available: <https://aclanthology.org/N18-1202>
- [10] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever *et al.*, "Improving language understanding by generative pre-training," 2018.
- [11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [12] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.
- [13] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," *arXiv preprint arXiv:1909.11942*, 2019.
- [14] Z. Zhang, Y. Wu, H. Zhao, Z. Li, S. Zhang, X. Zhou, and X. Zhou, "Semantics-aware bert for language understanding," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, 2020, pp. 9628–9635.
- [15] P. T. Hoang and P. X. Khoai, "Building a semantic role labelling toolkit for vietnamese," 2015.
- [16] T.-H. Pham, X.-K. Pham, and P. Le-Hong, "Building a semantic role labelling system for vietnamese," in *2015 Tenth International Conference on Digital Information Management (ICDIM)*. IEEE, 2015, pp. 77–84.
- [17] H. P. Le, T. H. Pham, X. K. Pham, T. M. H. Nguyen, T. L. Nguyen, and M. H. Nguyen, "Vietnamese semantic role labelling," 2018.
- [18] W. Lehnert, "The process of question answering. research report no. 88," 1977.
- [19] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "Squad: 100,000+ questions for machine comprehension of text," *arXiv preprint arXiv:1606.05250*, 2016.
- [20] K. Nguyen, V. Nguyen, A. Nguyen, and N. Nguyen, "A vietnamese dataset for evaluating machine reading comprehension," in *Proceedings of the 28th International Conference on Computational Linguistics*, 2020, pp. 2595–2605.
- [21] K. Van Nguyen, K. V. Tran, S. T. Luu, A. G.-T. Nguyen, and N. L.-T. Nguyen, "Enhancing lexical-based approach with external knowledge for vietnamese multiple-choice machine reading comprehension," *IEEE Access*, vol. 8, pp. 201 404–201 417, 2020.
- [22] K. Van Nguyen, T. Van Huynh, D.-V. Nguyen, A. G.-T. Nguyen, and N. L.-T. Nguyen, "New vietnamese corpus for machine reading comprehension of health news articles," *Transactions on Asian and Low-Resource Language Information Processing*, 2020.
- [23] R. Kadlec, M. Schmid, O. Bajgar, and J. Kleindienst, "Text understanding with the attention sum reader network," *arXiv preprint arXiv:1603.01547*, 2016.
- [24] W. Wang, N. Yang, F. Wei, B. Chang, and M. Zhou, "Gated self-matching networks for reading comprehension and question answering," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2017, pp. 189–198.
- [25] A. Trischler, T. Wang, X. Yuan, J. Harris, A. Sordoni, P. Bachman, and K. Suleman, "Newsqa: A machine comprehension dataset," *arXiv preprint arXiv:1611.09830*, 2016.
- [26] N. V. Kiet, T. Q. Son, N. T. Luan, H. V. Tin, L. T. Son, and N. L.-T. Ngan, "VLSP 2021 - ViMRC Challenge: Vietnamese Machine Reading Comprehension," *VNU Journal of Science: Computer Science and Communication Engineering*, vol. 38, no. 1, 2022.
- [27] X. Liu, W. Li, Y. Fang, A. Kim, K. Duh, and J. Gao, "Stochastic answer networks for squad 2.0," *arXiv preprint arXiv:1809.09194*, 2018.
- [28] Z. Li, S. He, H. Zhao, Y. Zhang, Z. Zhang, X. Zhou, and X. Zhou, "Dependency or span, end-to-end uniform semantic role labeling," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 6730–6737.
- [29] Z. Li, S. He, J. Cai, Z. Zhang, H. Zhao, G. Liu, L. Li, and L. Si, "A unified syntax-aware framework for semantic role labeling," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 2401–2411.
- [30] S. Narayanan and S. Harabagiu, "Question answering based on semantic structures," INTERNATIONAL COMPUTER SCIENCE INST BERKELEY CA, Tech. Rep., 2004.
- [31] G. Flieger, *Linguistically informed question answering*. German Research Center for Artificial Intelligence, Saarland University, 2007.
- [32] B. Ofoghi, J. Yearwood, and R. Ghosh, "A hybrid question answering schema using encapsulated semantics in lexical resources," in *Australasian Joint Conference on Artificial Intelligence*. Springer, 2006, pp. 1276–1280.
- [33] T. Mihaylov and A. Frank, "Discourse relation sense classification using cross-argument semantic similarity based on word embeddings," in *Proceedings of the CoNLL-16 shared task*, 2016, pp. 100–107.
- [34] C. Shi, S. Liu, S. Ren, S. Feng, M. Li, M. Zhou, X. Sun, and H. Wang, "Knowledge-based semantic embedding for machine translation," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016, pp. 2245–2254.
- [35] J. Tracey, S. Strassel, D. Graff, J. Wright, S. Chen, N. Ryant, S. Kulick, K. Griffith, D. Delgado, and M. Arrigo, "LORELEI Vietnamese Representative Language Pack," 2022. [Online]. Available: <https://hdl.handle.net/11272.1/AB2/JWPEIA>
- [36] K. Griffith, J. Tracey, A. Bies, and S. Strassel, "Simple semantic annotation and situation frames: Two approaches to basic text understanding in lorelei," in *LREC*, 2018.
- [37] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. [Online]. Available: <https://openreview.net/forum?id=Bkg6RiCqY7>