Proceedings of Machine Learning Research **318** (2026)
The 39th Canadian Conference on Artificial Intelligence
Canadian AI 2026
25-29 May 2026,Vancouver,BC

GlossAdapter: Enhancing Word Sense Disambiguation via LoRA Adapters

Vijayalakshmi Manikandan, Dunwei Wen*, M. Ali Akber Dewan
School of Computing and Information Systems, Athabasca University, Alberta, Canada

Abstract

Word sense disambiguation (WSD) is a long-standing problem in natural language processing (NLP). Recently, fine-tuned large pre-trained models with gloss and other lexical information have been used for WSD. But these models are parameter inefficient as the entire model needs to be trained. To deal with the problem, we propose GlossAdapter to WSD via Low-Rank Adaptation (LoRA) adapter modules and Part of Speech (POS) filtering. LoRA modules are parameter-efficient as they add only a few trainable parameters for a task, keeping the original weights of the pre-trained model frozen while maintaining the model quality. The proposed POS filtering aligns target word context with WordNet lexical categories to construct sentence-gloss pairs for effective model training. We fine-tune our model with SemCor3.0 dataset, and evaluated it with benchmark datasets Senseval-2, Senseval-3, SemEval-2013, and SemEval-2015. We perform experiments based on $\text{BERT}_{\text{base}}$ and $\text{RoBERTa}_{\text{large}}$ models. By adding only 0.5% of the parameters for $\text{RoBERTa}_{\text{large}}$, the results show that our LoRA adapter-based model combined with POS filtering outperforms the other state-of-the-art models.

Keywords: Word sense disambiguation, BERT, RoBERTa, low-rank adaptation, parameter-efficient fine-tuning.

1. Introduction

Word sense disambiguation (WSD) is a well-known long-standing problem in natural language processing (NLP), aiming to identify the meaning of a target word based on its context in a sentence. By ensuring the right sense of an ambiguous word is identified, WSD plays an important role in many NLP tasks such as question answering, sentiment analysis, information extraction etc., where accurate meaning and interpretation of ambiguous words are essential, and thus has many applications across various domains, such as law, bioinformatics, healthcare in identifying the correct information from a given data source [1].

Predefined word sense inventories such as WordNet [2] and BabelNet [3] contain a list of glosses for a word, which have been used by WSD to disambiguate the word meanings. Many researchers tried to perform WSD using various methods and approaches such as knowledge-based, hybrid, and recently deep learning-based methods involving model fine-tuning [4] [5]. While the deep fine-tuning methods can yield more promising results than traditional methods, they face various challenges including long training time and high demand for computing resources due to the full fine-tuning process, where all the original parameters are altered and the entire model is updated. The full fine-tuning may also increase the risk of overfitting, particularly on limited training data, lead to poor performance.

To address this, we develop a Low-Rank Adaptation(LoRA) [6] adapter-based approach, where new adapter modules are inserted in the transformer layers for WSD and trained while the original model weights remain frozen [7], which can reduce the training time and resource requirements while improving the model's accuracy and performance.

Our main contributions in this paper are as follows:

* corresponding_dunweiw@athabascau.ca

(1) We have developed an efficient and scalable approach for WSD, called GlossAdapter, by using a LoRA adapter-based approach along with the part-of-speech (POS) filtering. The resulting GlossAdapters are lightweight, task-specific modules that are easily transferable across different tasks and can be seamlessly integrated with other task-specific adapters, facilitating multi-task learning.
(2) Our model has outperformed current state of the art models such as $\text{EWISER}_{\text{hyper}}$ [8], KELESC [9], ConSeC [4], GlossBERT [10], $\text{RTWE}_{\text{large}}$ [11].
(3) The proposed GlossAdapter can significantly reduce the training time and memory footprint by training on only a tiny portion of parameters (0.5% of the parameters were trained for $\text{RoBERTa}_{\text{large}}$) by adopting the adapter-based method.

2. Methodology

Based on the state-of-the-art outlined in the Related Work (see Appendix B), in this section we explain our proposed GlossAdapter approach including LoRA-based parameter-efficient fine-tuning (PEFT) and POS filtering for achieving more efficient WSD. We also use $\text{BERT}_{\text{base}}$ and $\text{RoBERTa}_{\text{large}}$ as its foundation models, and SemCor3.0 as the input data to construct the training dataset with the sentence-gloss pairs.

2.1. LoRA Adapter based model training for WSD

As mentioned earlier, we have employed $\text{BERT}_{\text{base}}$ and $\text{RoBERTa}_{\text{large}}$ as the foundation models to develop the adapter for WSD problem. BERT [12] is a large language model using encoder-only bidirectional transformer-based architecture trained with several self-supervised learning methods, and capable of learning the semantic representations of a word by accounting for the bidirectional context. RoBERTa [13] is a BERT-based model trained with only masking-based method but on larger dataset.

Adapters are smaller modules which are added to the transformer layers and only the parameters of adapters are trained during fine-tuning [7]. We have used LoRA configuration [6] to develop and train our GlossAdapter model. The LoRA adapters are implemented via low-rank decompositions applied to the Query Q, Key K, and Value V projection matrices within the multi-head attention layers.

$$h = W_0 x + BAx \tag{2.1}$$

where h is the modified output, W_0 refers to pre-trained weight matrix, and $B \in R^{d\times r}$ and $A \in R^{r\times k}$ contain trainable parameters with the rank $r \ll \min(d,k)$ [6].

It has some key advantages: (i) a significant reduction in computational overhead and storage requirements; (ii) the preservation of the pre-trained weights, as it only requires updating a fraction of the total trainable parameters; (iii) faster training process compared to full-parameter fine-tuning, and (iv) the adapters can be switched dynamically during inference time that makes them transferable and used with other similar models. This allows the adapter modules to be shared across many other tasks.

2.2. POS filtering

Like many other WSD tasks, we use sentence-gloss pairs as input dataset for training the adapter-based model. The gloss is extracted from WordNet [2] and combined with the sentence for a given target word. Here is a detailed description of the training data construction, which includes our proposed POS filtering for reducing the ambiguity and noise in training data.

Developed using WordNet, the SemCor dataset contains sense keys (e.g., 'long%3:00:02::') with the POS information encoded in this structure. For instance, in the example sense key, the first digit '3' indicates that the POS category is adjective. Likewise, each POS category is mapped to a digit, which is used to extract the POS information. Based on this observation, we proposed a POS filtering for developing the sentence-gloss pairs from SemCor3.0. It involves identifying the target lemma and its associated POS category in the input dataset. By restricting WordNet gloss retrieval process to only including the specific lexical category of the target word, we were able to effectively prune the candidate space.

Thus, by filtering out irrelevant lexical categories, we ensure that the model's self-attention mechanism is dedicated to resolving intra-categorical nuances, which are often the most challenging to disambiguate. SemCor3.0 dataset is used in [10], but they have taken all possible glosses without filtering. Consequently, our approach has reduced the ambiguity and decreased the noise in the training dataset. In this way, a group of positive and negative examples are created using WordNet. The positive samples are assigned with label 1 and negative examples are assigned with label 0. This is done for all the target words in each sentence. Figure 1 shows the overall architecture of the proposed GlossAdapter, explaining how the input data passes through the various layers to give the output predictions.

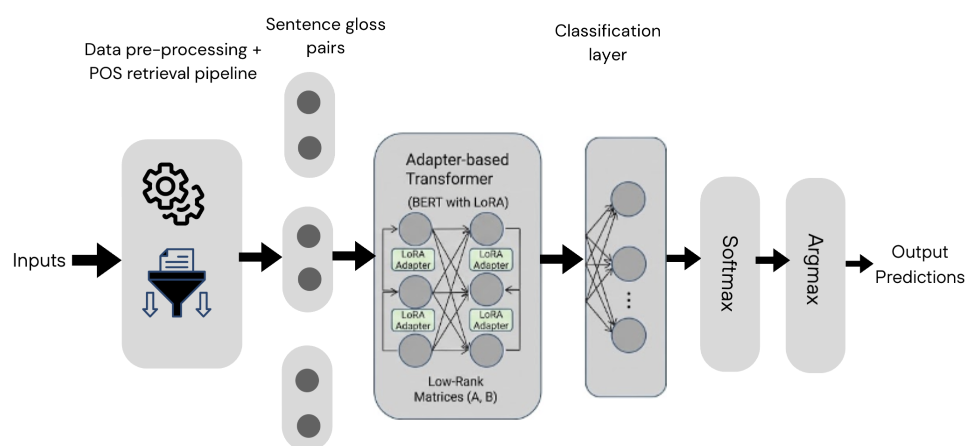

Figure 1. Overview of GlossAdapter architecture.

2.3. Evaluation

For the evaluation phase, the test datasets used include Senseval-2, Senseval-3, SemEval-2007, SemEval-2013, and SemEval-2015, and all test datasets combined to form an all-words dataset (ALL). Using the same approach introduced above, sentence-gloss pairs are constructed from the evaluation datasets without labels for the test dataset. The target words in the test datasets are surrounded by [TGT] and [/TGT] tags. The model outputs comprise of *logit* (raw similarity scores) and they are normalized to probabilities P_i using softmax function.

3. Experiments

3.1. Datasets

The training and test datasets used in this research are from Raganato et al. [14]. We used SemCor3.0 for fine-tuning the model. It is a well-recognized and widely used training dataset for WSD task as shown in [9] [8] [15]. It is a manually annotated corpus, containing

226,036 training instances including senses and glosses derived from WordNet 3.0. We have generated positive and negative samples for each target word in a sentence. The samples are restricted to POS that the target word belongs. Evaluation datasets from [14], including Senseval-2 [16], Senseval-3 [17], SemEval-2013 [18], SemEval-2015 [19] are used to evaluate the final model. These are the benchmark datasets obtained from the relevant Senseval and SemEval competitions. We have chosen SemEval-2007 as the development set similar to Huang et al. [10]. As mentioned above, a combination of them all test datasets (ALL) is also used in the evaluation.

3.2. Model Training

We have fine-tuned two main base models with LoRA adapter to identify the best performing GlossAdapter model, noted as LoRA-BERT$_{\text{base}}$ and LoRA-RoBERTa$_{\text{large}}$, where BERT$_{\text{base}}$ and RoBERTa$_{\text{large}}$ are used as the base models, respectively. The total number of parameters in pre-trained BERT$_{\text{base}}$ is 110M, with 12 layers of transformer blocks and 12 self-attention heads, out of which only 0.27% of parameters are trained using LoRA adapter. Whereas RoBERTa$_{\text{large}}$ has 355M parameters with 24 layers of transformer blocks and 16 self-attention heads, out of which only 0.5% of parameters are trained using LoRA adapter. For fine-tuning LoRA-BERT$_{\text{base}}$, we have used a learning rate of 3e-5 with a batch size of 64 and using 5 epochs for training. For LoRA-RoBERTa$_{\text{large}}$, we have used the same batch size and epochs but a different learning rate of 2e-5 because lower learning rate helps to stabilize training in case of larger models. The model is evaluated every 10,000 steps, and its checkpoints are saved. Finally, the best performing model is loaded from the saved checkpoints based on the evaluation metrics. Both models are trained on one A100 GPU. Following other existing works, we have used F1 score to evaluate and report the performance.

We used pre-defined PEFT framework from Hugging Face library for fine-tuning [20], with the following LoRA configuration parameters: the rank 'r' value as 8, alpha parameter for LoRA scaling value as 32 and the dropout probability for LoRA layers as 0.1.

3.3. Results and discussion

Table 1 compares our model's performance against the state-of-the-art results from recent studies on WSD. The comparison is done using development dataset SemEval-07 and test datasets Senseval-2, Senseval-3, SemEval-2013, SemEval-2015 as well as a combination of the test datasets ALL. F1 score metrics are calculated for the datasets.

The results show that our GlossAdapter model LoRA-RoBERTa$_{\text{large}}$ has outperformed all the other state-of-the-art models on Semeval-07, Senseval2, Senseval3, Semeval-2013 datasets developed on the same training set SemCor. RoBERTa$_{\text{large}}$ has 355 million parameters and using our proposed GlossAdapter we only need to train 0.5% of the parameters. Due to the significant reduction in the number of trainable parameters, we were able to complete the training in less than 4 hours using one A100 GPU.

The significant improvement can be attributed to our proposed GlossAdapter, which combines LoRA-based adapter structure with POS filtering. Firstly, we have used an LoRA adapter-based technique that helped us to train the model faster and utilize less compute resources. These adapters can be easily switched during inference time with other similar models which makes them easily transferable. Secondly, we developed a POS filtering that introduced a constrained data construction framework by filtering the POS of the target word and retrieving only the aligned WordNet glosses, thereby reducing the ambiguity and noise in training data. Also, it was preprocessed by adding [TGT] tags surrounding the target words and leveraged a suitable BERT based model for this task. This made an additional contribution to model performance improvement. Finally, the LoRA-based adapter configuration caused a substantial reduction in trainable parameter count, which

Model	**Training Dataset**	**SemEval -2007**	**Senseval -2**	**Senseval -3**	**SemEval -2013**	**SemEval -2015**	**ALL**
Current State-of-the-art Models							
GlossBERT [10]	SemCor	72.5	77.7	75.2	76.1	80.4	77.0
ESCHER [5]	SemCor	76.3	81.7	77.8	82.2	83.2	80.7
Rlarge [15]	SemCor	78.5	82.5	80.2	82.3	85.3	82.0
KELESC [9]	SemCor	76.7	82.2	78.1	82.2	83.0	81.2
$\text{EWISER}_{\text{hyper}}$ [8]	SemCor	75.2	80.8	79.0	80.7	81.8	80.1
ConSeC [4]	SemCor	77.4	82.3	79.9	83.2	85.2	82.0
$\text{RTWE}_{\text{large}}$ [11]	SemCor	74.5±0.4	83.4±0.2	82.9±0.8	82.1±0.2	85.3±0.3	82.7±0.1
GlossGPT [21]	SemCor	76.2	86.1	82.9	75.4	85.3	81.8
*Proposed **GlossAdapter** Models*							
$\text{LoRA-BERT}_{\text{base}}$	SemCor	90.0	83.6	80.3	84.5	86.2	82.8
LoRA-RoBERTa$_{\text{large}}$	**SemCor**	**91.3**	**87.7**	**85.1**	**88.8**	**89.2**	**87.2**

Table 1. F1 score % of WSD across test datasets and development dataset is reported here. The first block represents results of the current state of the art models from other research. The second block represents our model performance.

helped in significant savings in both computational expenditure and hardware resource utilization.

4. Conclusion

In this paper, we have proposed GlossAdapter, a novel approach to WSD using LoRA adapter-based models along with POS filtering. The experiments conducted show that our approach outperformed all the other state-of-the-art models on benchmark datasets SemEval-2007, Senseval-2, Senseval-3, and SemEval-2013 trained with the same dataset SemCor. Our ablation study (see Appendix A) provided additional clarity by isolating the performance impact of each constituent part and it confirms that the significant improvements are not the result of a single feature, but rather the collective contribution of the integrated LoRA adapters and POS filtering.

For future work, we will leverage more lexical information such as hypernym and hyponym example sentences to understand if it helps improve the model performance. We would also further enhance the performance of the proposed WSD approach and apply it in healthcare and education domains.

Acknowledgements

This research is part of Health Everywhere funded by the Major Innovation Fund program of the Ministry of Technology and Innovation, Government of Alberta (grant MIF-23-007).

References

[1] Y. Wang, K. Zheng, H. Xu, and Q. Mei. “Interactive Medical Word Sense Disambiguation Through Informed Learning”. In: *J. American Med. Info. Assoc.* 25.7 (2018), pp. 800–808.

[2] G. A. Miller. “WordNet: a lexical database for English”. In: *Communications of the ACM* 38.11 (1995), pp. 39–41.

[3] R. Navigli and S. P. Ponzetto. “BabelNet: Building a very large multilingual semantic network”. In: *Proc. of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*. 2010, pp. 216–225.

[4] E. Barba, L. Procopio, and R. Navigli. "ConSeC: Word sense disambiguation as continuous sense comprehension". In: *Proc. of the 2021 Conf. on Empirical Methods in Natural Language Processing (EMNLP)*. 2021, pp. 1492–1503.

[5] E. Barba, T. Pasini, and R. Navigli. "ESC: Redesigning WSD with extractive sense comprehension". In: *Proc. of the 2021 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*. 2021, pp. 4661–4672.

[6] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. "LoRA: Low-Rank Adaptation of Large Language Models". In: *Int. Conf. on Learning Representations (ICLR)*. 2022.

[7] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. De Laroussilhe, A. Gesmundo, M. Attariyan, and S. Gelly. "Parameter-Efficient Transfer Learning for NLP". In: *Proc. of the 36th Int. Conf. on Machine Learning*. 2019, pp. 2790–2799.

[8] M. Bevilacqua and R. Navigli. "Breaking through the 80% glass ceiling: Raising the state of the art in word sense disambiguation by incorporating knowledge graph information". In: *Proc. ACL*. 2020, pp. 2854–2864.

[9] G. Zhang, W. Lu, X. Peng, S. Wang, B. Kan, and R. Yu. "Word sense disambiguation with knowledge-enhanced and local self-attention-based extractive sense comprehension". In: *Proc. of the 29th Int. Conf. on Computational Linguistics (COLING)*. 2022, pp. 4061–4070.

[10] L. Huang, C. Sun, X. Qiu, and X.-J. Huang. "GlossBERT: BERT for word sense disambiguation with gloss knowledge". In: *Proc. EMNLP-IJCNLP*. 2019, pp. 3509–3514.

[11] X. Zhang, R. Zhang, X. Li, F. Kong, J. Chen, S. Mensah, and Y. Mao. "Word Sense Disambiguation by Refining Target Word Embedding". In: *Proc. of the ACM Web Conference 2023*. 2023, pp. 1405–1414.

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. "BERT: Pre-training of deep bidirectional transformers for language understanding". In: *Proc. NAACL-HLT vol. 1*. 2019, pp. 4171–4186.

[13] 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". In: *arXiv preprint arXiv:1907.11692* (2019).

[14] A. Raganato, J. Camacho-Collados, and R. Navigli. "Word sense disambiguation: A unified evaluation framework and empirical comparison". In: *Proc. of the 15th Conf. of the European Chapter of the Association for Computational Linguistics (EACL) Vol. 1*. 2017, pp. 99–110.

[15] Y. Song, X. C. Ong, H. T. Ng, and Q. Lin. "Improved word sense disambiguation with enhanced sense representations". In: *Proc. Findings of ACL: EMNLP*. 2021, pp. 4311–4320.

[16] P. Edmonds and S. Cotton. "Senseval-2: overview". In: *Proc. Senseval-2*. 2001, pp. 1–5.

[17] B. Snyder and M. Palmer. "The English all-words task". In: *Proc. Senseval-3*. 2004, pp. 41–43.

[18] R. Navigli, D. Jurgens, and D. Vannella. "SemEval-2013 task 12: Multilingual word sense disambiguation". In: *Second Joint Conf. on Lexical and Computational Semantics, Vol. 2: Proc. of the 7th Int. Workshop on Semantic Evaluation (SemEval)*. 2013, pp. 222–231.

[19] A. Moro and R. Navigli. "SemEval-2015 task 13: Multilingual all-words sense disambiguation and entity linking". In: *Proc. SemEval*. 2015, pp. 288–297.

[20] Hugging Face. *PEFT configuration and model*. Accessed: 2026-04-12. 2026. URL: https://huggingface.co/docs/peft/task_guides/lora_based_methods.

[21] D. Sumanathilaka, N. Micallef, and J. Hough. "GlossGPT: GPT for word sense disambiguation using few-shot chain-of-thought prompting". In: *Procedia Computer Science* 257 (2025), pp. 785–792.

[22] T. Blevins and L. Zettlemoyer. "Moving down the long tail of word sense disambiguation with gloss informed bi-encoders". In: *Proc. ACL*. 2020, pp. 1006–1017.

[23] J. Pfeiffer, I. Vulić, I. Gurevych, and S. Ruder. "MAD-X: An adapter-based framework for multi-task cross-lingual transfer". In: *Proc. EMNLP*. 2020, pp. 7654–7673.

[24] D. G. Olivares, L. Quijano, and F. Liberatore. "SANDWiCH: Semantical analysis of neighbours for disambiguating words in context ad Hoc". In: *Proc. NAACL-HLT Vol. 1*. 2025, pp. 7019–7033.

Appendix A. Ablation Study

We conducted ablation studies to assess and analyze the model's performance with and without LoRA adapters, and POS filtering. By removing the LoRA adapters and fine-tuning both the models $\text{BERT}_{\text{base}}$ and $\text{RoBERTa}_{\text{large}}$, we were able to see the impact of LoRA adapters on the model performance. Also, by removing the POS filtering, we were able to see its impact on the performance. Table A.1 shows the impact of integration of the LoRA adapter and using the POS filtering. Both LoRA adapter based models, LoRA-$\text{RoBERTa}_{\text{large}}$ and LoRA-$\text{BERT}_{\text{base}}$ outperform their fully fine-tuned version. This shows the impact of LoRA adapters on the model performance.

Our ablation studies have shown a critical trade-off between model performance and computational efficiency. While the fully fine-tuned variants demonstrated marginal performance gains over the baseline LoRA adapter (without POS filtering), these gains were accompanied with a significant increase in computational overhead. We argue that the benefits of LoRA adapter–specifically the reduction in number of trainable parameters (only 0.5% of parameters are trained for LoRA-$\text{RoBERTa}_{\text{large}}$ versus 100% of parameters trained in case of fully finetuned variants) and GPU hours–far outweigh the negligible accuracy delta, which saw a 86.2% for LoRA-$\text{BERT}_{\text{base}}$ versus 87% for fully fine-tuned $\text{BERT}_{\text{base}}$ for SemEval-2015 dataset. Combining LoRA + POS has shown better results. For instance, on Senseval-3, the LoRA-$\text{RoBERTa}_{\text{large}}$ + POS filtering achieved an accuracy of 85.1, a notable 2.1% increase over the 83.0 achieved by standard full fine-tuning $\text{RoBERTa}_{\text{large}}$.

Overall, the best performance was achieved by the LoRA-$\text{RoBERTa}_{\text{large}}$ + POS integrated model, which significantly outperformed all the other models while maintaining a low resource footprint. Hence, we conclude that our GlossAdapter, i.e., LoRA-based PEFT adapter with POS filtering, is the superior one for this task.

Ablation Study Models	**POS Filtering**	**Senseval -2**	**Senseval -3**	**SemEval -2013**	**SemEval -2015**
Base model $\text{BERT}_{\text{base}}$	No	51.2	43.7	52.8	53.9
Base model $\text{RoBERTa}_{\text{large}}$	No	53.7	45.7	52.4	52.4
Fully fined-tuned $\text{BERT}_{\text{base}}$	No	86.4	82.3	83.4	86.3
Fully fined-tuned $\text{RoBERTa}_{\text{large}}$	No	87.2	83.0	86.6	88.2
Fully fined-tuned $\text{BERT}_{\text{base}}$	Yes	86.1	83.0	84.4	87.0
Fully fined-tuned $\text{RoBERTa}_{\text{large}}$	Yes	86.0	83.6	87.0	87.7
LoRA-$\text{BERT}_{\text{base}}$	No	81.5	79	82.1	84.2
LoRA-$\text{RoBERTa}_{\text{large}}$	No	86.1	84.4	87.5	88.1
LoRA-$\text{BERT}_{\text{base}}$	Yes	83.6	80.3	84.5	86.2
LoRA-$\text{RoBERTa}_{\text{large}}$	**Yes**	**87.7**	**85.1**	**88.8**	**89.2**

Table A.1. Ablation study: F1 scores with and without LoRA, and POS filtering

Appendix B. Related Work

Researchers have developed different approaches to WSD. Barba et al. [5] proposed ESCHER to WSD by approaching it as a span extraction problem. They fed the transformer architecture with the input sentence including a target word, along with all its available sense definitions, to get the output of the start and end indices of the best suited gloss of the given target word. The model is trained using SemCor dataset as given by Raganato et al. [14] and tested on SemEval-2013 [18], SemEval-2015 [19] with promising results.

A similar approach is taken by Song et al. [15].They used not only the glosses but also example sentences, hypernyms, and other information from the lexical resources to train the model based on RoBERTa [13]. Training with SemCor plus Princeton WordNet Gloss Corpus has achieved F1 score of 82%. This has achieved better performance than ESCHERl [5].

Edoardo et al. [4] and Zhang et al. [9] approached WSD as text extraction problem and proposed ConSeC and KELESC, respectively. ConSeC utilizes both the context of a word and the senses definitions assigned to the nearby surrounding words and was evaluated using datasets in English and other languages, while KELESC feeds the transformer model with additional examples and definitions of related senses from the lexical resources, using a local self-attention transformer to reduce the computation burden due to the increased data.

Huang et al. [10] approached WSD as sentence-classification problem, and constructed context-gloss pairs using all possible senses from WordNet to train their model GlossBERT. They used SemCor3.0 for training dataset and used Senseval-2, Senseval-3, SemEval-2007, SemEval-2013, and SemEval-2015 for evaluation. GlossBERT is fine-tuned using the training data and outperforms other benchmark datasets [14].

Blevins et al. [22] targeted the problem of models becoming biased due to imbalanced annotated training data. They used a BERT-based bi-encoder model architecture composed of context encoder and gloss encoder, and the evaluation yielded a better performance than GlossBERT [10] on the same datasets.

Bevilacqua et al. [8] utilized the full capability of lexical knowledge information in neural architecture for WSD. They incorporated synset embeddings and proposed a neural architecture called EWISER. The experiment was conducted using a BERT model with different combinations of hypernym and hyponym and the F1 scores exceeded an 80% ceiling.

Pfeiffer et al. [23] proposed an adapter-based modular framework that has shown high performance in cross-lingual transfer across diverse languages using multilingual based model. In addition, they have developed task specific adapters assembled together with the language adapters for adapting a pretrained multilingual model to new languages.

Houlsby et al. [7] developed task adapter modules based on BERT for text classification tasks. Since fine-tuning a pre-trained model is a parameter inefficient, they leveraged adapters to solve the classification tasks on GLUE benchmark datasets. These adapter modules are easily transferable as they only add a few trainable parameters for any given tasks that help in achieving state-of-the-art performance.

Hu et al. [6] introduced Low-Rank Adaptation (LoRA), a PEFT strategy addressing the high computational cost of adapting Large Language Models. By freezing the pre-trained weights and injecting trainable rank-decomposition matrices into the Transformer layers, LoRA helps to significantly reduce the number of trainable parameters while maintaining the performance. Unlike traditional fine-tuning, LoRA represents weight updates through the product of two low-rank matrices - A and B.

Recently, Sumanathilaka et al. [21] employed a few-shot Chain-of-Thought (CoT) prompting strategy GlossGPT using GPT-4-Turbo for training, with WordNet glosses and synonyms in their prompts to enrich the context. Olivares et al. [24] proposed SANDWiCH incorporating a sophisticated semantic network refined from BabelNet using group algebra, yielding relatively higher performance with different but enhanced model and dataset including BabelNets.

Drawing inspiration from [6] [7] we have incorporated LoRA configuration in GlossAdapter. Inspired by [10], we also trained our models using sentence-gloss pairs; however, unlike their approach, we have not used all glosses from WordNet. We have developed POS filtering to align the target word's contextual usage with its corresponding entry in WordNet and restricted the gloss retrieval by excluding or filtering out synsets different from the lexical category. This ensures that the training dataset consists of only syntactically aligned definitions, hence reducing ambiguity during training and decreasing the noise in the dataset. Comparing our research with all these studies, none of them have leveraged PEFT along with POS filtering for WSD. Hence, our approach is novel in this way.