Instruction Tuning for Few-Shot Aspect-Based Sentiment Analysis

Siddharth Varia[1*] **Shuai Wang**[1*] **Kishaloy Halder**[1*] **Robert Vacareanu**[2†*]
Miguel Ballesteros[1] **Yassine Benajiba**[1] **Neha Anna John**[1]
Rishita Anubhai[1] **Smaranda Muresan**[1] **Dan Roth**[1]
[1]AWS AI Labs
[2]University of Arizona, Tucson, AZ, USA
{siddhvar, wshui, kishaloh, ballemig, benajiy, ranubhai, smaranm, drot}@amazon.com
rvacareanu@arizona.edu

Abstract

Aspect-based Sentiment Analysis (ABSA) is a fine-grained sentiment analysis task which involves four elements from user-generated texts: aspect term, aspect category, opinion term, and sentiment polarity. Most computational approaches focus on some of the ABSA sub-tasks such as tuple (aspect term, sentiment polarity) or triplet (aspect term, opinion term, sentiment polarity) extraction using either pipeline or joint modeling approaches. Recently, generative approaches have been proposed to extract all four elements as (one or more) quadruplets from text as a single task. In this work, we take a step further and propose a unified framework for solving ABSA, and the associated sub-tasks to improve the performance in few-shot scenarios. To this end, we fine-tune a T5 model with *instructional prompts* in a multi-task learning fashion covering all the sub-tasks, as well as the entire quadruple prediction task. In experiments with multiple benchmark data sets, we show that the proposed multi-task prompting approach brings performance boost (by absolute 6.75 F1) in the few-shot learning setting.

1 Introduction

Aspect-Based Sentiment Analysis (ABSA) is a fine-grained sentiment analysis task where the goal is to extract the sentiment associated with an entity and all its aspects (Liu, 2012; Pontiki et al., 2014, 2015, 2016; Schouten and Frasincar, 2015; Zhang et al., 2018; Nazir et al., 2020; Zhang et al., 2022). For example, in the context of Restaurant reviews the relevant aspects could be *food, ambience, location, service* with *general* used to represent the subject itself (i.e., restaurant). ABSA can provide valuable fine-grained information for businesses to analyze the aspects they care about. Annotated datasets have been released to foster research in this area (Pontiki et al., 2014, 2015, 2016).

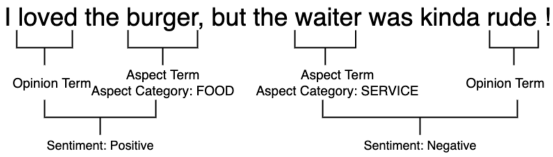

Figure 1: Illustrative orientation of four ABSA elements *i.e.,* Aspect Term, Aspect Category, Opinion Term, and Sentiment. The related tasks often involve predicting either everything together or a subset of them.

A full ABSA task aims to extract four elements from a user-generated text: aspect term, aspect category, opinion term and the sentiment polarity (see Figure 1 for an example). Most existing approaches have the focus on extracting some of these elements such as a single element (*e.g.,* aspect term), tuple (*e.g.,* aspect term, sentiment polarity), or triplet (*e.g.,* aspect term, aspect category, sentiment polarity) (Li et al., 2020; Hu et al., 2019; Xu et al., 2020). Recently, Zhang et al. (2021a) tackled the full ABSA task, under the name of Aspect Sentiment Quadruple Prediction (ASQP). Technically, most existing computational approaches have used extractive and discriminative models either in a pipeline or in an end-to-end framework (Wang et al., 2016; Yu et al., 2019; Cai et al., 2021) to address ABSA. Generative approaches have been recently shown to be effective for the full ABSA task and its sub-tasks (Zhang et al., 2021a,b; Yan et al., 2021). Most notably, Zhang et al. (2021a) used a sequence-to-sequence (seq-to-seq) model to address ASQP as a paraphrase generation problem. One important consideration is that modeling ABSA in a generative fashion allows for cross-task knowledge transfer.

*Indicates equal contribution.
†Work done during internship at AWS.

We go a step further and propose a unified model that can tackle multiple ABSA sub-tasks, including the ASQP task, and explore its effectiveness for low data scenarios. Recent work on large language models relies on the intuition that most natural language processing tasks can be described via natural language instructions and that models trained on these instructions show strong zero-shot performance on several tasks (Wei et al., 2021; Sanh et al., 2022). Based on this success, we propose a unified model based on multi-task prompting with instructional prompts using T5 (Raffel et al., 2020) to solve the full ABSA task *i.e.,* AQSP (Zhang et al., 2021a) and several of its associated sub-tasks addressed in the literature: 1) Aspect term Extraction (AE) (Jakob and Gurevych, 2010); 2) Aspect term Extraction and Sentiment Classification (AESC) (Yan et al., 2021); 3) Target Aspect Sentiment Detection (TASD), which aims to extract the aspect term, aspect category, and sentiment polarity (Wan et al., 2020); 4) Aspect Sentiment Triplet Extraction (ASTE), which aims to extract the aspect term, opinion term, sentiment polarity (Peng et al., 2020). We conduct an extensive set of experiments with multiple review datasets. Experimental results show that our proposed model achieves substantial improvement (6.75 F1 on average) against the state-of-the-art in few-shot learning scenario.

2 Methods

The four elements of ABSA form a quadruple as the sentiments are associated with both the aspect, and the opinion terms (*cf* Figure 1). In this work, we hypothesize that it is important to capture the interaction between these components not only at the quadruple level, but also within a subset of these four elements.

We consider multiple factorized sub-tasks involving one or more of the four elements to be predicted. We pose it as a combination of five Question Answering (QA) tasks as illustrated in Figure 2. For each QA task, an instructional prompt is used to train a seq-to-seq model to learn one or more ABSA elements – referred to as Instruction Tuning (IT). Our formulation enables learning all sub-tasks via Multi-Task Learning (MTL).

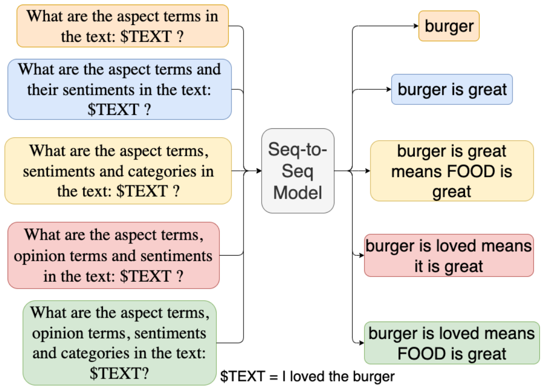

Figure 2: Instruction tuning to solve the sub-tasks related to ABSA. We devise multiple prompts to instruct a seq-to-seq model to learn in multi-task learning manner.

2.1 Input Transformation

First, we transform each sentence in the corpus using the instruction templates provided for each task provided in Table 1. To make the model learn the underlying tasks, we use multiple paraphrased instruction templates[1] for a task, and sample randomly when preparing a batch during training the seq-to-seq model. However, the target output sequence remains unchanged irrespective of the template sampled for a task.

2.2 Model Training

Next, we perform IT with the seq-to-seq model. We train it in a MTL fashion where input-output combinations are sampled from all tasks simultaneously. We use the following loss for model training:

$$\mathcal{L} = -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n}\log p_\theta(y_i|y_1,\ldots,y_{i-1},\mathbf{x_t}). \quad (1)$$

where $\mathbf{x_t}$ is the transformed input sequence ($\mathbf{x}$) for t^{th} task. θ is the set of model parameters. n is the length of output sequence. y_i is the i^{th} token in output sequence. T is the number of tasks. The model parameters are updated using Adam optimizer with weight decay (Loshchilov and Hutter, 2019).

2.3 Output Transformation

Finally, we transform the output using the templates provided in the rightmost column in Table 1.

[1]Complete list of instructions in Appendix

Task	\$AT	\$AC	\$S	\$OT	Input Instruction	Output
Aspect Extraction (AE)	✓				Given the text: \$TEXT, what are the aspect terms in it ? What are the aspect terms in the text: \$TEXT ?	Template: \$AT Literal: *burger*
Aspect term Extraction and Sentiment Classification (AESC)	✓		✓		Given the text: \$TEXT, what are the aspect terms and their sentiments ? What are the aspect terms and their sentiments in the text: \$TEXT ?	Template: \$AT is \$S Literal: *burger is great*
Target Aspect Sentiment Detection (TASD)	✓	✓	✓		Given the text: \$TEXT, what are the aspect terms, sentiments and categories ? What are the aspect terms, sentiments and categories in the text: \$TEXT ?	Template: \$AT is \$S means \$AC is \$S Literal: *burger is great means food is great*
Aspect Sentiment Triplet Extraction (ASTE)	✓		✓	✓	Given the text: \$TEXT, what are the aspect terms, opinion terms and sentiments ? What are the aspect terms, opinion terms and sentiments in the text: \$TEXT ?	Template: \$AT is \$OT means it is \$S Literal: *burger is loved means it is great*
Aspect Sentiment Quadruple Prediction (ASQP)	✓	✓	✓	✓	Given the text: \$TEXT, what are the aspect terms, opinion terms, sentiments and categories ? What are the aspect terms, opinion terms, sentiments and categories in the text: \$TEXT ?	Template: \$AT is \$OT means \$AC is \$S Literal: *burger is loved means food is great*

Table 1: The factorized sub-tasks in ABSA. Each of them covers a sub-set of all four prediction targets. \$AT: Aspect Term; \$AC: Aspect Category; \$S: Sentiment; \$OT: Opinion Term; \$TEXT: input text. Both templates and literal values (for \$TEXT = *I loved the burger*) are shown for Output against each task.

In case there is more than one quadruple in the output, we use a special separation token `[SSEP]`. We map sentiment classes positive, negative and neutral to *great, bad* and *ok* respectively in the output similar to (Zhang et al., 2021a). During inference, we apply the reverse transformations to recover the quadruples for evaluation.

3 Experiments

As this work is one of the first few attempts towards studying few-shot learning in ABSA context, unsurprisingly, there is a lack of standard few-shot datasets. We emulate few-shot data drawing inspiration from the literature (Halder et al., 2020; Ma et al., 2022) for our experiments.

3.1 Datasets: Few-shot Preparation

We use three datasets, REST15, REST16 and LAPTOP14 from (Zhang et al., 2021a). For the first two, we shuffle the data with fixed random seed, and select first few samples so that there are at least k samples from each aspect category.[2] As LAPTOP14 does not have aspect category annotations, we select k examples per sentiment class instead, following the same principle (statistics in Table 5).

[2]Not feasible to guarantee exactly k samples since an example can have multiple aspect categories. (Ma et al., 2022)

3.2 Baselines and Models for Comparison

As a strong baseline, we consider PARAPHRASE (or PARA) model[3] – the current state-of-the-art for TASD, ASTE, and ASQP tasks (Zhang et al., 2021a). It uses the same backbone model as of ours, which ensures comparable results, and there is no influence of increased model size.

To understand the impact of all components in our approach, we consider three model ablations:

1. **Text**: \$TEXT is directly used as input
2. **IT**: \$TEXT is transformed to instructions
3. **IT-MTL**: IT with MTL covering all tasks[4]

3.3 Experimental Setup

We use `t5-base` (Raffel et al., 2020) as the backbone for our models. Results are averaged over 5 runs with random seeds (*cf* Section A.2 for all details). Micro F1 is the evaluation metric following previous work (Zhang et al., 2021a).

3.4 Results

We present results for all the datasets in Table 2. Since, LAPTOP14 lacks aspect category annotations, TASD and ASQP are not applicable. We make four key observations from the results.
Ablation Study: First, IT beats Text in most settings proving effectiveness of our instructions. Second, we observe that IT-MTL outperforms others

[3]Other competitive models can be found in (Zhang et al., 2021a). Since PARA has outperformed them, we focus on it.
[4]*cf* Table 4 in Appendix for worked out example

on REST15, and REST16 substantially in few-shot settings, except on LAPTOP14 as IT-MTL underperforms on AE task. This might be attributed to the absence of TASD, ASQP tasks. Overall, we observe the trend IT-MTL > IT > Text.

Baseline Comparison: Third, our proposed IT-MTL approach outperforms PARA[5] comfortably in most few-shot settings across all datasets with a performance boost of 6.75 F1 on average. We observe some exceptions in case of LAPTOP14, where PARA yields highest score, closely followed by IT on AE task due to the missing tasks that require aspect category annotation.

Forth, upon training on full training data (*Full*), IT-MTL achieves comparable (2 out of 7 cases) or better (4 out of 7 cases) F1 against PARA. Only exception is ASTE on LAPTOP14 dataset, where IT-MTL falls behind by $\sim$ 2 F1 points. Overall, we conclude that in few-shot settings, our proposed IT-MTL leverages on the context from multiple tasks, and improves the generalization of the underlying seq-to-seq model across all the ABSA tasks.

4 Conclusion

In this paper, we posed ABSA as an instruction tuning based seq-to-seq modeling task. We factorized the overall quadruple prediction task into multiple sub-tasks. We proposed a multi-task learning based approach using a pre-trained seq-to-seq model. We experimented with customer reviews from two domains and showed that our approach gives superior performance compared to baseline models in few-shot and full fine-tuning scenarios.

Task	Model	K=5	K=10	K=20	K=50	Full
AE	Text	43.95	54.38	59.75	61.75	72.53
	IT	**45.24**	55.1	60.33	64.15	72.73
	IT-MTL	44.18	**56.57**	**62.65**	**67.22**	**74.03**
AESC	Text	37.33	47.68	50.6	56.69	66.2
	IT	**39.4**	**49.43**	52.06	58.4	66.49
	IT-MTL	38.99	47.62	**53.58**	**59.54**	**67.06**
TASD	PARA.	22.13	33.74	39.44	45.16	**60.07**
	Text	22.55	36.37	42.28	48.52	58.93
	IT	22.92	36.52	43.2	50.14	59.46
	IT-MTL	**27.05**	**36.81**	**43.56**	**50.24**	59.76
ASTE	PARA.	19.66	25.89	30.56	33.30	**52.67**
	Text	18.49	30.17	35.66	41.49	51.13
	IT	22.38	32.11	36.67	41.65	51.7
	IT-MTL	**22.7**	**33.52**	**37.78**	**43.84**	52.6
ASQP	PARA.	11.9	20.18	23.92	27.86	46.91
	Text	12.15	22.19	28.82	33.96	46.79
	IT	13.3	24.35	29.66	36.78	46.59
	IT-MTL	**15.54**	**25.46**	**31.47**	**37.72**	**46.99**

(a) REST15

Task	Model	K=5	K=10	K=20	K=50	Full
AE	Text	52.7	58.5	61.49	67.21	78.57
	IT	55.64	59.36	63.75	68.14	79.09
	IT-MTL	**59.41**	**61.87**	**66.88**	**71.18**	**79.41**
AESC	Text	49.13	54.54	57.05	62.75	73.55
	IT	51.93	55.29	59.96	63.45	**74.32**
	IT-MTL	**52.42**	**55.37**	**60.22**	**65.14**	74.07
TASD	PARA.	27.48	37.07	44.85	49.72	**68.14**
	Text	30.65	38.39	46.72	54.04	67.12
	IT	34.38	38.58	47.66	55.16	67.52
	IT-MTL	**40.45**	**42.41**	**48.83**	**55.82**	67.73
ASTE	PARA.	27.84	32.39	36.18	41.72	62.09
	Text	28.44	38.23	42.12	50.9	62.68
	IT	33.08	**41.12**	44.08	51.69	**63.49**
	IT-MTL	**35.75**	38.95	**44.75**	**52.94**	62.27
ASQP	PARA.	16.61	23.68	29.14	36.09	**57.90**
	Text	20.98	28.06	35.04	45.26	57.41
	IT	23.86	30.02	37.2	46.9	57.48
	IT-MTL	**27.02**	**31.66**	**38.06**	**47.48**	57.61

(b) REST16

Task	Model	K=5	K=10	K=20	K=50	Full
AE	Text	**34.64**	42.26	51.11	59.62	75.43
	IT	34.29	**47.4**	52.39	**63.86**	76.11
	IT-MTL	31.54	42.73	**53.08**	63.71	**76.93**
AESC	Text	21.68	30.7	37.74	50.39	**66.76**
	IT	23.28	**36.55**	43.39	52.92	65.97
	IT-MTL	**25.01**	34.44	**44.5**	**53.75**	66.07
ASTE	PARA.	**14.33**	22.00	27.71	33.40	**60.70**
	Text	10.1	16.27	26.37	39.65	58.69
	IT	12.6	21.31	30.03	41.91	60.14
	IT-MTL	14.18	**24.09**	**32.39**	**42.62**	58.74

(c) LAPTOP14

Table 2: Results on (a) REST15, (b) REST16, (c) LAPTOP14 datasets for few(k)-shot learning, and full dataset. Best is **bolded**, second-best is underlined. Our proposed MTL outperforms PARA, and other ablations comfortably in most cases in few-shot scenarios.

References

Hongjie Cai, Rui Xia, and Jianfei Yu. 2021. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 340–350, Online. Association for Computational Linguistics.

Kishaloy Halder, Alan Akbik, Josip Krapac, and Roland Vollgraf. 2020. Task aware representation of sentences for generic text classification. In *COLING 2020, 28th International Conference on Computational Linguistics*.

[5] Results are populated with code released by authors.

Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, and Yiwei Lv. 2019. Open-domain targeted sentiment analysis via span-based extraction and classification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 537–546.

Niklas Jakob and Iryna Gurevych. 2010. Extracting opinion targets in a single and cross-domain setting with conditional random fields. In *Proceedings of the 2010 conference on empirical methods in natural language processing*, pages 1035–1045.

Kun Li, Chengbo Chen, Xiaojun Quan, Qing Ling, and Yan Song. 2020. Conditional augmentation for aspect term extraction via masked sequence-to-sequence generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7056–7066.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *ICLR*.

Jie Ma, Miguel Ballesteros, Srikanth Doss, Rishita Anubhai, Sunil Mallya, Yaser Al-Onaizan, and Dan Roth. 2022. Label semantics for few shot named entity recognition. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1956–1971.

Ambreen Nazir, Yuan Rao, Lianwei Wu, and Ling Sun. 2020. Issues and challenges of aspect-based sentiment analysis: a comprehensive survey. *IEEE Transactions on Affective Computing*.

Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2020. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):8600–8607.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, San Diego, California. Association for Computational Linguistics.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. SemEval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495, Denver, Colorado. Association for Computational Linguistics.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Victor Sanh, Albert Webson, Colin Raffel, Stephen Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven Le Scao, Arun Raja, et al. 2022. Multitask prompted training enables zero-shot task generalization. In *The Tenth International Conference on Learning Representations*.

Kim Schouten and Flavius Frasincar. 2015. Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 28(3):813–830.

Hai Wan, Yufei Yang, Jianfeng Du, Yanan Liu, Kunxun Qi, and Jeff Z. Pan. 2020. Target-aspect-sentiment joint detection for aspect-based sentiment analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):9122–9129.

Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2016. Recursive neural conditional random fields for aspect-based sentiment analysis. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 616–626.

Jason Wei, Maarten Bosma, Vincent Y Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. 2021. Finetuned language models are zero-shot learners. *arXiv preprint arXiv:2109.01652*.

Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020. Position-aware tagging for aspect sentiment triplet extraction. *arXiv preprint arXiv:2010.02609*.

Hang Yan, Junqi Dai, Tuo Ji, Xipeng Qiu, and Zheng Zhang. 2021. A unified generative framework for aspect-based sentiment analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2416–2429, Online. Association for Computational Linguistics.

Jianfei Yu, Jing Jiang, and Rui Xia. 2019. Global inference for aspect and opinion terms co-extraction based on multi-task neural networks. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, 27(1):168–177.

Lei Zhang, Shuai Wang, and Bing Liu. 2018. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1253.

Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, and Wai Lam. 2021a. Aspect sentiment quad prediction as paraphrase generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9209–9219, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2021b. Towards generative aspect-based sentiment analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 504–510, Online. Association for Computational Linguistics.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2022. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *arXiv preprint arXiv:2203.01054*.

A Appendix

A.1 List of input instruction prompts

All the input instruction prompts are presented in table 3. During training, we randomly select the input prompt for all tasks at the start of each epoch. During evaluation, we again select random input prompt for each example in the test set for all the tasks.

A.2 Hyperparameters

We set the learning rate to 3e-4 for all the experiments in this paper. We train each model for a fixed number of 20 epochs similar to Zhang et al.. For full-shot experiments, we use a batch size of 16. For k=5, 10, 20 and 50 we use a batch size of 2, 2, 4 and 8 respectively. The maximum sequence length is set to 160. Longer sequences are truncated and shorter sequences are padded. Finally, we use Adam optimizer with weight decay.

A.3 Dataset Statistics

Table 5 presents the number of sentences in each dataset. Please note that for LAPTOP14 dataset, the few-shot data for different values of K was selected based on sentiment classes instead of Aspect category due to lack of category annotations.

Task	Input Prompts
AE	Given the text: $TEXT, what are the aspect terms in it ?
	What are the aspect terms in the text: $TEXT ?
ASE	Given the text: $TEXT, what are the aspect terms and their sentiments ?
	What are the aspect terms and their sentiments in the text: $TEXT ?
TASD	Given the text: $TEXT, what are the aspect terms, sentiments and categories ?
	What are the aspect terms, sentiments and categories in the text: $TEXT ?
	Given the text: $TEXT, what are the aspect terms, categories and sentiments ?
	What are the aspect terms, categories and sentiments in the text: $TEXT ?
ASTE	Given the text: $TEXT, what are the aspect terms, opinion terms and sentiments ?
	What are the aspect terms, opinion terms and sentiments in the text: $TEXT ?
	Given the text: $TEXT, what are the opinion terms, aspect terms and sentiments ?
	What are the opinion terms, aspect terms and sentiments in the text: $TEXT ?
ASQP	Given the text: $TEXT, what are the aspect terms, opinion terms, sentiments and categories ?
	What are the aspect terms, opinion terms, sentiments and categories in the text: $TEXT ?
	Given the text: $TEXT, what are the aspect terms, opinion terms, categories and sentiments ?
	What are the aspect terms, opinion terms, categories and sentiments in the text: $TEXT ?
	Given the text: $TEXT, what are the opinion terms, aspect terms, sentiments and categories ?
	What are the opinion terms, aspect terms, sentiments and categories in the text: $TEXT ?
	Given the text: $TEXT, what are the opinion terms, aspect terms, categories and sentiments ?
	What are the opinion terms, aspect terms, categories and sentiments in the text: $TEXT ?

Table 3: List of input instruction prompts for all the five sub-tasks. $TEXT is the place holder for actual text.

Ablation	Input Prompt
Text	$TEXT
IT	What are the aspect terms in the text: $TEXT?
IT-MTL	What are the aspect terms in the text: $TEXT?
	What are the aspect terms and their sentiments in the text: $TEXT?
	Given the text: $TEXT, what are the aspect terms, sentiments and categories?
	Given the text: $TEXT, what are the aspect terms, opinion terms and sentiments?
	What are the aspect terms, opinion terms, sentiments and categories in the text: $TEXT ?

Table 4: Illustration of input prompts to the seq-to-seq model for various ablations of our proposed approach.

	Rest15					Rest16					Laptop14				
	K=5	K=10	K=20	K=50	Full	K=5	K=10	K=20	K=50	Full	K=5	K=10	K=20	K=50	Full
Train	25	46	86	181	834	22	43	77	179	1264	11	19	40	106	906
Dev	21	35	68	140	209	26	42	73	159	316	8	16	34	86	219
Test	537					544					328				

Table 5: Number of sentences in each dataset. The same test set was used for few-shot and full-shot evaluation.