Generative Aspect-Based Sentiment Analysis with Contrastive Learning and Expressive Structure

Joseph J. Peper and **Lu Wang**
Computer Science and Engineering
University of Michigan
Ann Arbor, MI
{jpeper, wangluxy}@umich.edu

Abstract

Generative models have demonstrated impressive results on Aspect-based Sentiment Analysis (ABSA) tasks, particularly for the emerging task of extracting Aspect-Category-Opinion-Sentiment (ACOS) quadruples. However, these models struggle with *implicit* sentiment expressions, which are commonly observed in opinionated content such as online reviews. In this work, we introduce **GEN-SCL-NAT**, which consists of two techniques for improved structured generation for ACOS quadruple extraction. First, we propose **GEN-SCL**, a supervised contrastive learning objective that aids quadruple prediction by encouraging the model to produce input representations that are discriminable across key input attributes, such as sentiment polarity and the existence of implicit opinions and aspects. Second, we introduce **GEN-NAT**, a new structured generation format that better adapts pre-trained autoregressive encoder-decoder models to extract quadruples in a generative fashion.

Experimental results show that GEN-SCL-NAT achieves top performance across three ACOS datasets, averaging 1.48% F1 improvement, with a maximum 1.73% increase on the LAPTOP-L1 dataset. Additionally, we see significant gains on implicit aspect and opinion splits that have been shown as challenging for existing ACOS approaches.

1 Introduction

Aspect-based sentiment analysis (ABSA) is the task of extracting fine-grained sentiment information from text. Applications of this task range from social media opinion mining and empathetic dialog assistants to product review analysis (Zhang et al., 2022).

Common ABSA subtasks involve identifying *aspect term* mentions and their corresponding aspect *categories*, associating them with supporting *opinion terms*, and/or *sentiment polarity* that is implicitly or explicitly expressed within the text.

While over 30% of sentiment expressions contain implicit language (Cai et al., 2021), existing methods struggle with these, performing significantly worse on examples with implicit aspects and/or opinions. These cases, such as "it took an hour to be seated", are more difficult for models as they lack common indicative *explicit* aspect and opinion terms, e.g., "the service" and "slow".

In this work, we address the ACOS quadruple extraction task, particularly for formulations supporting implicit aspects and opinions. We build off Zhang et al. (2021), a state-of-the-art T5-based technique outputting parseable structured ACOS quadruple predictions for a given text. We introduce **GEN-SCL-NAT**[1], consisting of two novel modifications to this approach. First, we modify the model training objective, adding **GEN-SCL**, an auxiliary supervised contrastive learning objective that tasks the model to discriminably represent examples across key characteristics such as the existence of implicit aspects and opinions, and the expressed sentiment polarity. A supervised contrastive loss is used to maximize the margin between inconsistent examples and minimize it between consistent examples. Second, we introduce a modified generation target format with **GEN-NAT**, where we improve the naturalness of the target format in three ways: by (1) replacing existing aspect category labels with human-readable category descriptions, (2) enforcing a reproducible "scan-based" inter-quadruple ordering for multi-quad cases, and (3) modifying the intra-quadruple format to reflect a natural causal ordering of quadruple components that aligns with auto-regressive decoding. Figure 1 outlines the ACOS task and our GEN-SCL-NAT approach.

Leveraging these techniques, we show improved performance on ACOS benchmarks, with respective ACOS F1 improvements of 1.64%, 1.08% and

---

[1] Code and models are made available at https://github.com/jpeper/GEN_SCL_NAT.
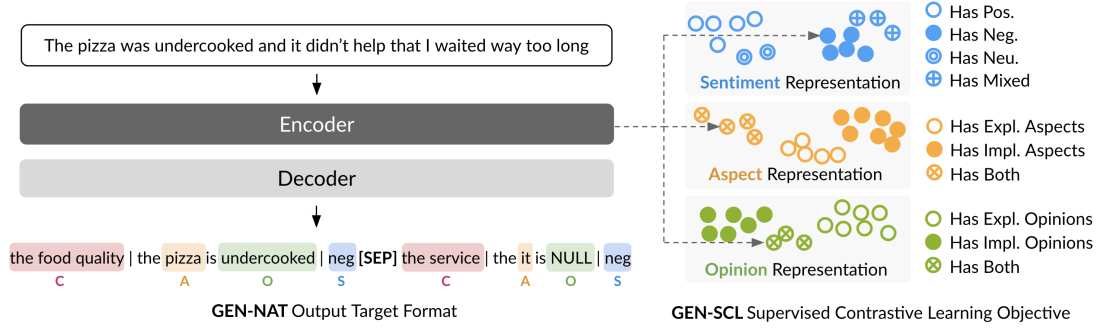
Figure 1: Overview of the GEN-SCL-NAT ACOS approach. GEN-NAT uses natural category descriptions and intuitive quadruple ordering to align the target output with the encoder-decoder's pre-training and auto-regressive nature. GEN-SCL trains the model to discriminably represent three sentiment characteristics during training: (1) sentiment polarity, (2) aspect term type, and (3) opinion term type. For each characteristic, we project the summed encoder final hidden states into a 1024-dimension characteristic representation via a simple fully-connected layer.

1.73% on the REST, LAPTOP, and LAPTOP-L1 datasets. We also see significant gains on examples with challenging implicit language, averaging 1.47% improvement on this subset. Finally, we conduct an ablation study that shows the complementary behavior of GEN-NAT and GEN-SCL.

2 ACOS Quadruple Extraction Task Formulation

We follow Zhang et al. (2021) and Cai et al. (2021) in formulating ABSA as a joint quadruple extraction task, where the goal is to extract an unordered set of ACOS quadruples $Q_1, Q_2, ...Q_n$ in text T, where $Q_i = (a_i, c_i, o_i, s_i)$ contains a corresponding aspect term, aspect category, opinion term, and sentiment polarity. Quadruples can lack clear supporting aspect and/or opinion spans, and these cases are marked as *implicit*. Figure 1 displays the ACOS task and quadruple components.

3 Methodology

In this section we (1) introduce *GEN-SCL*, a task-specific supervised contrastive learning (SCL) objective, and (2) propose *GEN-NAT*, an enhanced ACOS quadruple generation target format that addresses weaknesses in existing methods.

3.1 GEN-SCL Supervised Contrastive Loss

We propose GEN-SCL, an auxiliary SCL objective that encourages the encoder-decoder model to discriminably represent several key characteristics of the input while concurrently fine-tuning for the downstream generation task of ACOS quadruple extraction. We task the model with learning representations of example-level **Sentiment**, **Aspect** and **Opinion** characteristics. Figure 1 indicates the label sets for each characteristic.

Representation Generation Process Similar to Li et al. (2021), we generate representations for example x_i by feeding the sum-pooled encoder representation $Mean(Encode(x_i))$ through a single unique fully-connected layer FC_c for each characteristic $c \in \{Sentiment, Aspect, Opinion\}$. This generates representation $\mathbf{h}_{ci}$. In our experiments the dimensionality of the input and outputs to the fully-connected layers are both 1024.

SCL Formulation Supervised contrastive learning encourages the model to maximize the representation similarity between *same-label* examples, and to minimize it for *different-label* examples. We follow Sedghamiz et al. (2021) in their general SCL formulation, where for characteristic c and training example x_i in training mini-batch M, the loss is:

$$\mathcal{L}_i^c = \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{e^{\text{sim}(\mathbf{h}_{ci}, \mathbf{h}_{cp})/\tau}}{\sum_{b \in B(i)} e^{\text{sim}(\mathbf{h}_{ci}, \mathbf{h}_{cb})/\tau}} \quad (1)$$

To ensure each example $x_i \in M$ has at least one *same-label* example for comparison, we first extend M with one dropout-altered view per mini-batch element, perturbing each example representation $\mathbf{h_{ci}}$ with dropout probability $p = 0.1$ while maintaining the original label. This becomes M^*. $B(i) \equiv M^* \setminus x_i$ is all other examples in the extended mini-batch, and $P(i) \equiv \{p \in B(i) : y_{cp} = y_{ci}\}$ is the subset with a matching label.

Final training objective We add to the existing decoder cross-entropy loss $\mathcal{L}_{CE}$ our three characteristic-specific losses:

$$\mathcal{L}_{total} = \mathcal{L}_{CE} + \alpha_1 \mathcal{L}_{SCL\_sentiment} \quad (2)$$
$$+ \alpha_2 \mathcal{L}_{SCL\_aspect} + \alpha_3 \mathcal{L}_{SCL\_opinion}$$

In our experiments we set $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$, tuning α and the SCL temperature τ on the dev set, with values reported in Table 5.

3.2 GEN-NAT Structured Generation Format

Existing Structured Generation Formulation
We build off Zhang et al. (2021) who define the generation target as a linearization of $Q_i = (a_i, c_i, o_i, s_i)$ to an output format P:

$$P_c(c_i) \; is \; P_s(s_i) \; because \; P_a(a_i) \; is \; P_o(o_i) \qquad (3)$$

$P_c(c_i) = c_i$ and $P_s(s_i)$ is the mapping of s_i options [positive, neutral, negative] to [*"great", "okay", "bad"*]. $P_a(a_i) = a_i$ and $P_o(o_i) = o_i$ for explicit cases, but are respectively "it" and "null" for implicit cases. Quadruples are concatenated with a separator token to form the final output:

$$P(Q_1) \; [SSEP] \; ... \; [SSEP] \; P(Q_n) \qquad (4)$$

GEN-NAT Structured Generation Modifications
We implement three modifications to the existing paraphrase generation format to aid in decoding. (1) We revise $P_c(c_i)$, now defining it as a mapping of the raw c_i (e.g. "Laptop#Usability") to a natural category description (e.g. "the laptop usability"). Note: we also change the sentiment linearization $P_s(s_i)$, simply using [*"positive", "neutral", "negative"*], as we observe *"okay"* can often imply negative sentiment within review text (e.g "The food was just okay. I wouldn't return."). (2) For multi-output cases, per Meng et al. (2020), a consistent ordering is useful in training even when predicting unordered sets; following them, we use a "scan-based" ordering, outputting quadruples by their last-occurring explicit aspect or opinion term. Quadruples with only implicit aspect and opinion spans are generated last in random order. (3) We revise the quadruple linearization format, following Mao et al. (2022) in outputting the quadruple elements in a natural top-down causal ordering: (c, a, o, s). Now, sentiment prediction is conditioned on aspect and opinion outputs. We also partition some quadruple components with "|" to reduce parsing ambiguity when mapping the predictions back to ACOS format.

Our resultant NAT generation format is as follows:

$$P_c(c_i) \mid the \; P_a(a_i) \; is \; P_o(o_i) \mid P_s(s_i) \qquad (5)$$

4 Experiment Setup

We detail the experiment setup for evaluating our techniques on the ACOS task.

	REST	LAPTOP	LAPTOP-L1
#Categories	13	121	21
#Sentences	2286	4076	4076
#EAEO Quads	2429 (66.4%)	3269 (56.8%)	3269 (56.8%)
#IAEO Quads	530 (14.5%)	910 (15.8%)	910 (15.8%)
#EAIO Quads	350 (9.57%)	1237 (21.5%)	1237 (21.5%)
#IAIO Quads	349 (9.54%)	342 (5.94%)	342 (5.94%)
#Quads/Sent	1.60	1.42	1.42

Table 1: Dataset statistics. Over 33% of REST quadruples contain implicit language, as do 43% in LAPTOP*. E: explicit, I: implicit, A: aspect, O: opinion. E.g., *IAEO* refers to "implicit aspect, explicit opinion".

	REST	LAPTOP	LAPTOP-L1
BERT Backbone			
TAS-BERT-ACOS*	33.53	27.31	–
Extract-Classify-ACOS*	44.61	35.80	–
T5 Backbone			
Seq2Path*	58.06	41.45	–
PARAPHRASE	60.97	44.08	60.73
GEN-NAT-SCL (ours)	62.62†	45.16†	62.46†

Table 2: Overall F1 performance of quadruple extraction techniques on the REST, LAPTOP and LAPTOP-L1 datasets. Scores are averaged over 5 unique runs. *Results are from Cai et al. (2021) and Mao et al. (2022). †: method is significantly better than PARAPHRASE (one-tailed unpaired t-test, $p < 0.05$).

ACOS Datasets Table 1 reports dataset statistics. The RESTaurant and LAPTOP datasets (Cai et al., 2021) are drawn from restaurant and e-commerce domains and are ACOS-labeled reviews including implicit aspects and opinions. LAPTOP-L1 differs from LAPTOP only by the category label granularity, using only the 21 top-level categories of the two-level category hierarchy in LAPTOP.

Model Comparisons We compare five ACOS techniques, considering both implicit aspect and implicit opinion cases as done in Cai et al. (2021).

• **TAS-BERT-ACOS**: Cai et al. (2021) devise a two-step pipelined method, incorporating TAS-BERT (Wan et al., 2020) for triplet extraction.

• **Extract-Classify-ACOS**: Cai et al. (2021) leverage BERT (Devlin et al., 2018) to extract aspect-opinion pairs then perform category and sentiment linking.

• **Seq2Path**: Mao et al. (2022) generate ACOS quadruples as paths of a tree, supporting multiple quadruples through multi-beam search and filtering candidates via a learned discriminator token.

• **PARAPHRASE**: A generative T5 structured paraphrase generation model, producing a parseable ACOS sequence prediction (Zhang et al., 2021).

• **GEN-SCL-NAT**: Finally, our proposed method, including the SCL and NAT components.

	REST				LAPTOP				LAPTOP-L1			
Method	EAEO	IAEO	EAIO	IAIO	EAEO	IAEO	EAIO	IAIO	EAEO	IAEO	EAIO	IAIO
TAS-BERT-ACOS	33.6	31.8	14.0	39.8	26.1	41.5	10.9	21.2	-	-	-	-
Extract-Classify-ACOS	45.0	34.7	23.9	33.7	35.4	39.0	16.8	18.6	-	-	-	-
PARAPHRASE	65.4	53.3	45.6	49.2	45.7	51.0	33.0	**39.6**	64.2	65.2	49.3	**53.8**
GEN-SCL-NAT (ours)	**66.5**	**56.5**†	**46.2**	**50.7**	**45.8**	**54.0**†	**34.3**	39.6	**65.6**	**66.7**	**51.5**†	53.7

Table 3: Breakdown of F1 performance per example split, with each split comprising reviews containing that quadruple type. E: explicit, I: implicit, A: aspect, O: opinion. $\dagger$: method is significantly better than PARAPHRASE (one-tailed unpaired t-test, $p < 0.05$).

	REST	LAPTOP	LAPTOP-L1	Avg. Δ
PARAPHRASE	60.97	44.08	60.73	-1.49
GEN-SCL-NAT	**62.62**†	45.16†	**62.46**†	-
(-Sentiment SCL)	62.36†	**45.72**†	62.03†	-0.05
(-Aspect SCL)	61.78	45.14†	61.68†	-0.55
(-Opinion SCL)	60.68	44.71	61.77†	-1.03
(-All SCL)	62.18†	44.03	62.12†	-0.64
(-Multi-quad Ordering)	59.92	44.63	61.00	-1.56
(-Natural Category Labels)	61.28	43.94	61.96†	-1.02
(-Intra-quad Ordering)	61.39	45.43†	61.48†	-0.65
(-All NAT)	59.89	43.86	61.36	-1.26

Table 4: Ablation analysis of the GEN-SCL-NAT model. F1 scores are reported, averaged over 5 runs. We compare to the baseline PARAPHRASE model. The best result is bolded. $\dagger$: method is significantly better than PARAPHRASE (one-tailed unpaired t-test, $p < 0.05$).

Experiment Details We report scores averaged over five runs each with different random seeds. For our GEN-* models, we adopt the 770M parameter T5-large (Raffel et al., 2019) as our pre-trained generative encoder-decoder model. We report further experiment details and hyperparameter settings in Appendix A.

5 Results

We evaluate on the task of exact quadruple extraction using the F1 metric, where a correct extraction requires all components are correct.

Overall Performance Table 2 reports the overall performance on the ACOS task. We find that GEN-SCL-NAT outperforms other approaches on all three datasets. The BERT pipeline approaches struggle heavily, likely due to error-accumulation over several sub-tasks and limited pre-training alignment. Among T5 methods, Seq2Path performs worst by over 3% on the two reported datasets, perhaps due to limitations in their beam-search candidate pruning method; in contrast PARAPHRASE and our approach generate all quadruples in one output sequence.

Explicit vs Implicit Performance Breakdown Table 3 reports the results on four implicit/explicit aspect/opinion data splits. Notably, our method is adept at the IAEO and EAIO subsets, with average 2.56% and 1.37% respective increases of F1

over PARAPHRASE. We see *IO splits are consistently more challenging than *EO splits, as models have no supervised signal with which to localize implicit opinions in the text. However, we still see consistent EAIO gains from our method, along with IAIO gains of 1.5% for the REST dataset. These results speak to the improved predictive power of our techniques—we consistently outperform on the difficult and sparse implicit subsets while still increasing the average EAEO performance by 0.88%.

6 Additional Analyses

6.1 GEN-SCL-NAT Ablation Study

We ablate the GEN-SCL-NAT model by withholding various components of the technique, including each of the NAT enhancements and SCL losses. Table 4 reports ablation results.

SCL Ablation We ablate each of the three SCL objectives. We withhold each component of the SCL loss, finding the opinion loss most impactful (average decrease of 1.03%), and also observe decreases from the aspect and sentiment ablations. Notably, we see the losses considering implicit language phenomena (*Aspect SCL*, *Opinion SCL*) are the most impactful, indicating the benefits of modeling these challenging examples. We additionally withhold all SCL losses (*All SCL*), observing noticeable consistent performance degradation on all datasets with the SCL component removed.

NAT Ablation Of the three NAT components, we see that excluding scan-based multi-quad ordering (outputting quadruples in the order they are mentioned in the text) has the largest impact on overall performance. Notably, the impact is largest on the REST dataset ($62.62 \rightarrow 59.92$) which contains more quadruples per example than the Laptop datasets (average of 1.60 vs. 1.42). This validates our intuition that proper output ordering is significant for multi-quadruple examples. Next, we see that natural category labels are impactful, especially for the LAPTOP dataset ($45.16 \rightarrow 43.94$) containing a much larger cate-
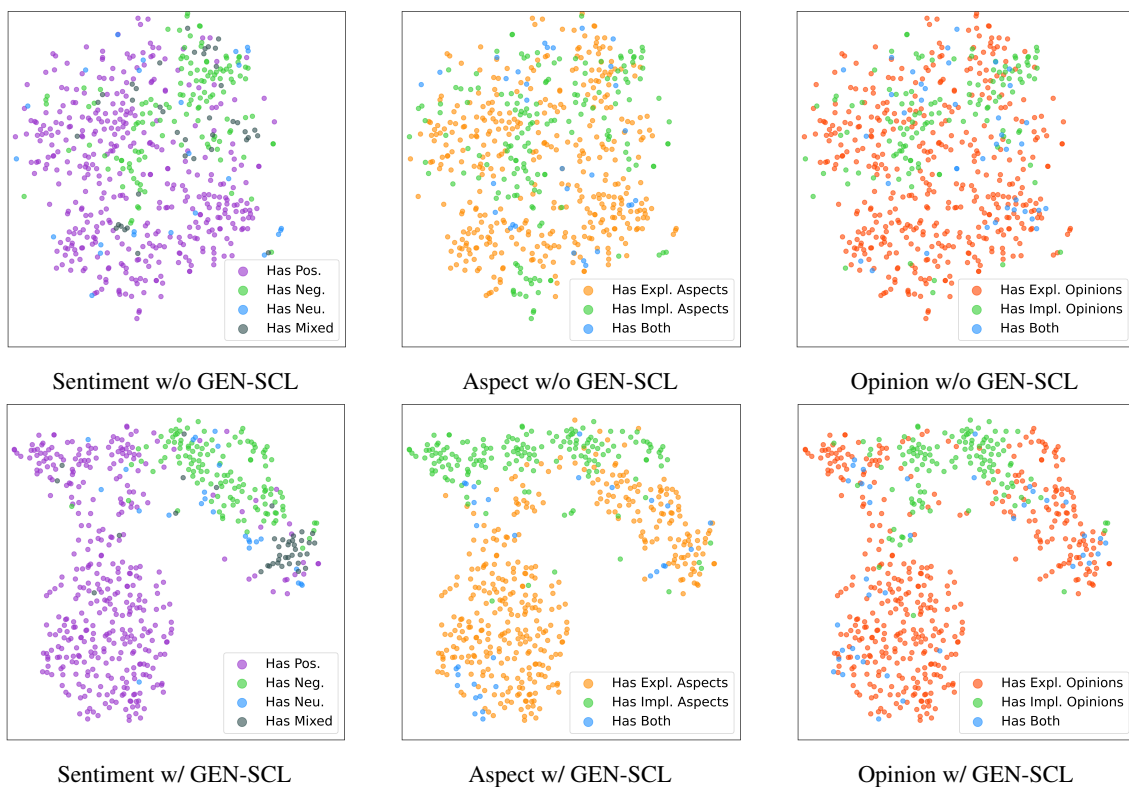
Figure 2: T-SNE visualization of mean-pooled encoder final layer on the Restaurant dataset. Our GEN-SCL objective encourages the encoder to produce with discriminable representations of three key input characteristics.

gory labelset (121 labels versus 13 and 21 for the Restaurant and Laptop-L1 datasets). Finally, we see GEN-SCL-NAT benefits from the intra-quad ordering, although not as noticeably as the other two improvements. We see improvements on REST and LAPTOP-L1, while performance was comparable with and without for LAPTOP ($45.16 \rightarrow 45.43$). Overall, our full GEN-SCL-NAT method yields consistently strong performance, with its components working in unison to address challenging properties that arise in the ACOS task (implicit aspects and opinions, multi-quad examples, large category labelsets).

6.2 GEN-SCL Representations

To better understand the behavior of the GEN-SCL objective on the model hidden representations, we generate t-SNE (Maaten and Hinton, 2008) visualizations of the mean-pooled final encoder layer. Figure 2 displays results on the Restaurant test set. We see GEN-SCL enables the encoder to simultaneously represent sentiment, aspect, and opinion information effectively.

7 Related Work

ACOS quadruple prediction is an emerging ABSA task with structured generation techniques producing top results on the ACOS task (Zhang et al.,

2021; Mao et al., 2022; Zhang et al., 2022). Meng et al. (2020) empirically demonstrate the importance of output ordering and formatting in generative unordered set prediction techniques for keyphrase generation, motivating our exploration of this direction. Supervised contrastive learning (Khosla et al., 2020) is a popular technique for representation learning and has proven useful for NLP tasks (Sedghamiz et al., 2021). Li et al. (2021) apply it to ABSA tasks during pre-training, but only for representing positive vs negative sentiment. Our novel methods combine generation format improvements with a task-specific supervised contrastive learning objective that learns to represent key ACOS characteristics.

8 Conclusion

In this work, we introduce GEN-NAT, a modified ACOS generation output format encompassing three methods for improving the naturalness of the decoded output sequences. We combine this with GEN-SCL, our novel task-specific application of supervised contrastive learning that learns improved example representations leading to downstream gains. Our proposed GEN-SCL-NAT method demonstrates state-of-the-art results, both overall and for challenging implicit-sentiment splits.

Acknowledgements

Limitations

While effective on concise sentence-level tasks, we have not yet explored ACOS structured generation for document-level and/or sentiment-dense inputs. Approaches such as ours and Zhang et al. (2021) generate the prediction as a single output, and we may encounter issues such as output structure validity when handling longer examples with higher quadruple frequency. Our evaluation of GEN-SCL is also confined to generative sequence prediction models. While it may generalize to other formulations, we have not yet explored this direction.

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.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. 2020. Supervised contrastive learning. *Advances in Neural Information Processing Systems*, 33:18661–18673.

Zhengyan Li, Yicheng Zou, Chong Zhang, Qi Zhang, and Zhongyu Wei. 2021. Learning implicit sentiment in aspect-based sentiment analysis with supervised contrastive pre-training. *arXiv preprint arXiv:2111.02194*.

Laurens Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605.

Yue Mao, Yi Shen, Jingchao Yang, Xiaoying Zhu, and Longjun Cai. 2022. Seq2path: Generating sentiment tuples as paths of a tree. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 2215–2225.

Rui Meng, Xingdi Yuan, Tong Wang, Sanqiang Zhao, Adam Trischler, and Daqing He. 2020. An empirical study on neural keyphrase generation.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683*.

Hooman Sedghamiz, Shivam Raval, Enrico Santus, Tuka Alhanai, and Mohammad Ghassemi. 2021. Supcl-seq: Supervised contrastive learning for downstream optimized sequence representations. *arXiv preprint arXiv:2109.07424*.

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. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9122–9129.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and Jamie Brew. 2019. Huggingface's transformers: State-of-the-art natural language processing. *CoRR*, abs/1910.03771.

Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, and Wai Lam. 2021. 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. 2022. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *arXiv preprint arXiv:2203.01054*.

A   Implementation Details

We use Huggingface Transformers (Wolf et al., 2019) for training our T5 models. We conduct our experiments on the Nvidia A40 GPU with 48GB VRAM.

Training   We largely use the T5 parameters specified in Zhang et al. (2021), making the following adjustments after performing manual accuracy-based hyperparameter tuning on the provided validation sets: We train for 45 epochs using batch size of 32, and a learning rate of 9e-5 for all experiments. For GEN-SCL, we tune α, the SCL loss weighting factor, and the τ temperature parameter per dataset, with these values reported in Table 5. We use the standard dataset train/validation/test splits provided by the authors of the REST and LAPTOP* datasets. Training takes approximately 1 hour for the RESTAURANT dataset, and 1.5 hours for the LAPTOP* datasets.

	α	τ
REST	0.05	0.25
LAPTOP	0.05	0.25
LAPTOP-L1	0.005	0.25

Table 5: We report the parameters used in the GEN-SCL supervised contrastive learning objective. α is the loss weighting factor, and τ is the temperature value that determines how severely to punish hard negative examples.

Dataset	Raw Category Label	GEN-NAT Category Label
REST	LOCATION#GENERAL	the location
	FOOD#PRICES	the food prices
	FOOD#QUALITY	the food quality
LAPTOP	OS#GENERAL	the operating system overall
	OS#DESIGN_FEATURES	the operating system features
	HARD_DISC#PRICE	the hard drive price
LAPTOP-L1	OS	the operating system
	HARD_DISC	the hard drive

Table 6: Examples of the aspect categories used for the ACOS task. Our GEN-NAT technique better leverages T5 pre-training by using human-readable descriptive category labels.

Prediction We use beam search to decode the ACOS output sequence; we set beam size to 5. Prediction on the RESTAURANT and LAPTOP test sets takes 5 minutes with an evaluation batch size of 32.

B GEN-NAT Category Mappings

One component of our GEN-NAT approach consists of replacing existing raw aspect category labels with human-readable natural category descriptions. Table 6 provides examples of these reformatted GEN-NAT category labels.