

# Few-shot Cross-lingual Aspect-Based Sentiment Analysis with Sequence-to-Sequence Models

Jakub Šmíd<sup>1,2</sup>[0000-0002-4492-5481], Pavel Přibáň<sup>1</sup>[0000-0002-8744-8726], and Pavel Král<sup>1,2</sup>[0000-0002-3096-675X]

<sup>1</sup> University of West Bohemia in Pilsen

Faculty of Applied Sciences, Department of Computer Science and Engineering

<sup>2</sup> NTIS – New Technologies for the Information Society

Univerzitni 27328, 301 00 Plzeň, Czech Republic

{jaksmid,pribanp,pkral}@kiv.zcu.cz

<https://nlp.kiv.zcu.cz>

**Abstract.** Aspect-based sentiment analysis (ABSA) has received substantial attention in English, yet challenges remain for low-resource languages due to the scarcity of labelled data. Current cross-lingual ABSA approaches often rely on external translation tools and overlook the potential benefits of incorporating a small number of target language examples into training. In this paper, we evaluate the effect of adding few-shot target language examples to the training set across four ABSA tasks, six target languages, and two sequence-to-sequence models. We show that adding as few as ten target language examples significantly improves performance over zero-shot settings and achieves a similar effect to constrained decoding in reducing prediction errors. Furthermore, we demonstrate that combining 1,000 target language examples with English data can even surpass monolingual baselines. These findings offer practical insights for improving cross-lingual ABSA in low-resource and domain-specific settings, as obtaining ten high-quality annotated examples is both feasible and highly effective.

**Keywords:** Cross-lingual aspect-based sentiment analysis · Aspect-based sentiment analysis · Sentiment analysis · Transformers

## 1 Introduction

Aspect-based sentiment analysis (ABSA) is a fine-grained sentiment analysis task that goes beyond assigning an overall sentiment label to a piece of text. Instead, it identifies specific opinion targets – such as products, services, or their attributes – and determines the sentiment expressed toward each of them. To capture this information, ABSA typically models three sentiment elements [13]: aspect term ( $a$ ), aspect category ( $c$ ), and sentiment polarity ( $p$ ). In the sentence “*The staff was helpful*”, the elements correspond to “*staff*”, “*service*”, and “*positive*”, respectively. Some inputs include implicit aspect terms, as in “*Delightful experience!*”, where the target is not stated directly and is often labelled as “*NULL*”.

ABSA tasks vary in complexity based on predicted elements and joint extraction. Early work focused on simpler tasks like aspect term extraction and sentiment polarity classification. Recent work has shifted towards compound tasks that jointly predict multiple sentiment elements, such as aspect category sentiment analysis (ACSA) [11], end-to-end ABSA (E2E-ABSA) [19], aspect category term extraction (ACTE) [10], and target-aspect-sentiment detection (TASD) [18]. Table 1 presents selected ABSA task output formats.

Table 1: Outputs of selected ABSA tasks for input: “*Great soup, expensive coffee*”.

Task	Output	Example output
ACSA	$\{(c, p)\}$	$\{(\text{food}, \text{POS}), (\text{drinks}, \text{NEG})\}$
E2E-ABSA	$\{(a, p)\}$	$\{(\text{“soup”}, \text{POS}), (\text{“coffee”}, \text{NEG})\}$
ACTE	$\{(a, c)\}$	$\{(\text{“soup”}, \text{food}), (\text{“coffee”}, \text{drinks})\}$
TASD	$\{(a, c, p)\}$	$\{(\text{“soup”}, \text{food}, \text{POS}), (\text{“coffee”}, \text{drinks}, \text{NEG})\}$

Although ABSA has been widely studied in English, multilingual support is essential for real-world applications, where annotating target-language data is costly. Cross-lingual ABSA transfers knowledge from high-resource languages like English to low-resource ones. Recent work [6, 7, 22] combines multilingual models like XLM-R [3] with machine translation to generate pseudo-labelled data, but this often fails to capture language-specific nuances in user-generated texts. It is also less effective for complex tasks like TASD, which require structured predictions. Sequence-to-sequence models with constrained decoding offer a more effective alternative [15].

While most cross-lingual ABSA studies focus on zero-shot transfer, few-shot examples – such as a small number of high-quality target language annotations – can be much more resource-efficient. However, the impact of such examples remains underexplored. In this work, we address these gaps by investigating the effects of adding a limited number of labelled target-language examples on compound cross-lingual ABSA tasks.

Our main contributions include: 1) A comprehensive evaluation of four ABSA tasks across six target languages using two state-of-the-art sequence-to-sequence models, including a previously unexplored task and target language in cross-lingual ABSA. 2) Demonstrating that even ten target language examples lead to substantial performance improvements, reducing the need for techniques like constrained decoding. 3) Showing that few-shot models can outperform monolingual baselines with sufficient target language supervision. 4) An error analysis highlighting the most challenging elements of sentiment prediction, offering insights for future improvements.

## 2 Related Work

Early cross-lingual ABSA research [1, 4, 5] focused on simple tasks involving a single sentiment element using machine translation or cross-lingual embeddings.

More recent work [6–8, 22] targets E2E-ABSA with machine translation and multi-lingual encoder-only Transformer models like XLM-R [3], enhanced by techniques like distillation [22], contrastive learning [7], or dynamic loss weighting [8].

Sequence-to-sequence models like mT5 [21] and mBART [17] have been applied to cross-lingual ABSA with constrained decoding as an alternative to machine translation [15]. These models handle compound tasks such as E2E-ABSA, ACTE, and TASD, demonstrating flexibility in capturing complex sentiment structures. Fine-tuned large language models are also explored [15], extending earlier work in monolingual English ABSA [14].

### 3 Methodology

In this section, we describe our method for tackling the triplet task (TASD), which can be readily adapted to tuple-based tasks with minimal modifications. Figure 1 provides an overview of the proposed approach.

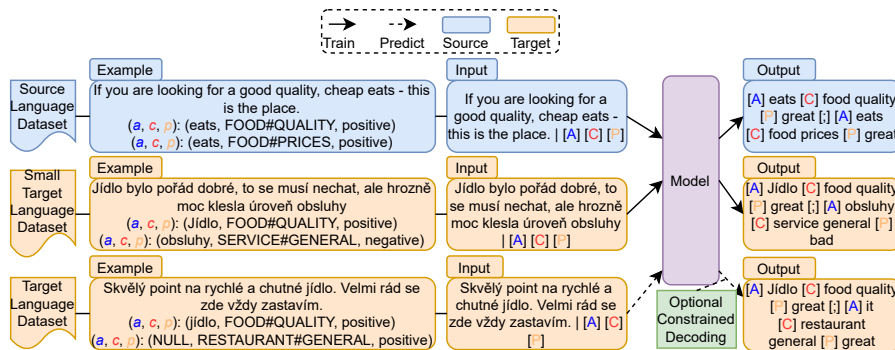


Fig. 1: Overview of the training approach. The method involves converting input labels into natural language phrases, fine-tuning on source language data alongside a few target language examples, and generating predictions on target language inputs. Constrained decoding is optionally used to improve output quality.

#### 3.1 Problem Definition

Given an input sentence, the goal is to predict all sentiment tuples  $T = (a, c, p)$ , where each tuple consists of an aspect term ( $a$ ), aspect category ( $c$ ), and sentiment polarity ( $p$ ). Following previous work [15] for a fair comparison, we convert each element ( $a, c, p$ ) into a natural language representation ( $e_a, e_c, e_p$ ), with a few examples illustrated in Figure 1. For example, we map implicit aspect terms to “it”, “positive” sentiment polarity to “great”, and “negative” sentiment polarity to “bad”, while keeping explicit aspect terms and aspect categories in their original form.

### 3.2 Input and Output Construction

To build input and output sequences for the model, we follow the strategy proposed in [15], using special markers to denote each sentiment element: [A] for  $e_a$ , [C] for  $e_c$ , and [P] for  $e_p$ . These markers are used to prefix each element in the target sequence and are also appended to the input sequence to guide the model’s decoding process. For instance, given the sentence “*The staff was very helpful*” yield the output: “[A] *staff* [C] *service general* [P] *great*”. If a sentence contains multiple sentiment tuples, we concatenate their representations using the [·] separator. Figure 1 includes several examples of such input-output pairs.

### 3.3 Constrained Decoding

Constrained decoding (CD) [2], shown in [15] to be effective for cross-lingual ABSA with sequence-to-sequence models, mitigates errors where a model trained on a source language (e.g. English) generates aspect terms in the source language instead of target one (e.g. Dutch). CD limits token generation based on the input and prior outputs. For example, when generating an aspect category, only tokens from the set of valid categories are allowed. For aspect terms, generation is limited to tokens from the original sentence and “*it*” for implicit aspects. We use the same CD method as in [15] and evaluate whether few-shot target language examples can offer similar benefits, potentially reducing the need for CD.

### 3.4 Training

We fine-tune a pre-trained sequence-to-sequence (encoder-decoder) model on constructed input–output pairs. The encoder processes the input sequence  $x$  into a contextual representation  $\mathbf{e}$ , and the decoder generates the output  $y$  by modelling the probability  $P_{\Theta}(y|\mathbf{e})$ , where  $\Theta$  denotes the model parameters. During training, each token  $y_i$  is predicted based on  $\mathbf{e}$  and previously generated tokens. The model is optimized by minimizing the negative log-likelihood loss over a target sequence of length  $n$ :

$$\mathcal{L} = - \sum_{i=1}^n \log p_{\Theta}(y_i | \mathbf{e}, y_{<i}). \quad (1)$$

## 4 Experimental Setup

We conduct experiments on four ABSA tasks: ACSA, E2E-ABSA, ACTE, and T ASD. To the best of our knowledge, this paper is the first to examine ACSA in cross-lingual settings.

We conduct experiments using the SemEval-2016 dataset [10] with restaurant reviews in six languages: English (en), Spanish (es), French (fr), Dutch (nl), Russian (ru), and Turkish (tr), with official training and test splits. A validation set is created by splitting the original training data in a 9:1 ratio, as in [15]. Additionally, we use the CsRest-M dataset [16], containing real-world restaurant

reviews in Czech (cs), which is divided into training, validation, and test sets. Czech has not been explored in cross-lingual ABSA before. Table 2 summarizes dataset statistics for each language.

Table 2: Dataset statistics for each language.

		Cs	En	Es	Fr	Nl	Ru	Tr
Train	Sentences	2,151	1,800	1,863	1,559	1,549	3,289	1,108
	Triplets	4,386	2,266	2,455	2,276	1,676	3,697	1,386
Dev	Sentences	240	200	207	174	173	366	124
	Triplets	483	241	265	254	184	392	149
Test	Sentences	798	676	881	694	575	1,209	144
	Triplets	1,609	859	1,072	954	613	1,300	159

#### 4.1 Experimental Details

We employ multilingual sequence-to-sequence models, mT5 [21] and mBART [17], accessed via the HuggingFace Transformers library [20]. For all experiments, each model is fine-tuned for 20 epochs with a batch size of 16, using greedy decoding during inference. We use Adafactor [12] optimizer with a learning rate of 1e-4 for mT5 and AdamW [9] with a learning rate of 1e-5 for mBART. Hyperparameters were selected for stable validation performance across languages and tasks. Experiments run on an NVIDIA L40 GPU with 48 GB of memory.

We use English as the source language and the other languages as target languages for all experiments and perform model selection based on performance on the English validation set. For few-shot experiments, we select the first  $n$  examples from the target language’s training set. The data are not ordered by label or difficulty, and the initial subset reflects the overall label distribution, ensuring representative coverage. This approach also ensures consistency across runs and models.

#### 4.2 Evaluation Metrics & Compared Methods

We report results averaged over five runs with different random seeds and 95% confidence intervals. The primary evaluation metric is the micro F1-score, standard in ABSA research, where a predicted sentiment tuple is considered correct only if all components exactly match the gold standard.

We compare our few-shot cross-lingual results with the zero-shot baselines reported in [15], where applicable – that is, for all languages except Czech and all tasks except ACSA. Zero-shot results for Czech and ACSA, as well as all few-shot results, are newly reported in this paper. Using the same models and data splits ensures a fair evaluation of the few-shot effect.

## 5 Results

Table 3 presents the results for the four ABSA tasks using mT5, comparing different numbers of few-shot examples and monolingual training. As expected, monolingual training consistently yields the best performance across all tasks and languages, highlighting the advantage of abundant in-language supervision.

Table 3: Cross-lingual micro F1-scores on four tasks in six target languages using mT5, compared to monolingual (mono) performance, with varying few-shot examples (FS). **Bold** marks better constrained decoding (CD) results with non-overlapping 95% confidence intervals. Asterisked (\*) results are from [15].

Task Setup FS		Without constrained decoding						With constrained decoding						
		Cs	Es	Fr	Nl	Ru	Tr	Cs	Es	Fr	Nl	Ru	Tr	
ACSA	Mono	76.6 <sup>±0.2</sup>	77.1 <sup>±0.2</sup>	69.2 <sup>±0.8</sup>	74.1 <sup>±0.3</sup>	78.0 <sup>±0.8</sup>	74.2 <sup>±1.5</sup>	76.5 <sup>±1.1</sup>	77.4 <sup>±0.6</sup>	69.0 <sup>±0.7</sup>	74.3 <sup>±0.9</sup>	77.7 <sup>±0.5</sup>	74.4 <sup>±4.6</sup>	
	Cross-lingual	0	68.0 <sup>±0.5</sup>	71.3 <sup>±0.9</sup>	64.5 <sup>±1.3</sup>	69.6 <sup>±0.4</sup>	72.5 <sup>±1.0</sup>	71.3 <sup>±1.6</sup>	67.6 <sup>±0.3</sup>	71.7 <sup>±0.9</sup>	63.3 <sup>±0.5</sup>	69.8 <sup>±1.1</sup>	72.1 <sup>±0.3</sup>	72.3 <sup>±1.6</sup>
		2	67.2 <sup>±0.7</sup>	72.1 <sup>±0.7</sup>	63.8 <sup>±0.9</sup>	69.9 <sup>±0.6</sup>	71.7 <sup>±0.6</sup>	70.0 <sup>±1.5</sup>	67.5 <sup>±0.2</sup>	71.4 <sup>±0.1</sup>	64.4 <sup>±1.3</sup>	70.0 <sup>±1.3</sup>	72.2 <sup>±0.8</sup>	71.9 <sup>±1.5</sup>
		5	66.8 <sup>±1.2</sup>	71.1 <sup>±0.5</sup>	63.4 <sup>±1.1</sup>	69.2 <sup>±0.4</sup>	72.3 <sup>±0.7</sup>	69.5 <sup>±1.8</sup>	66.5 <sup>±0.9</sup>	71.8 <sup>±0.5</sup>	63.8 <sup>±1.3</sup>	69.0 <sup>±1.4</sup>	72.8 <sup>±0.6</sup>	70.7 <sup>±1.4</sup>
		10	68.5 <sup>±0.7</sup>	71.2 <sup>±0.8</sup>	64.4 <sup>±1.2</sup>	71.2 <sup>±1.3</sup>	72.0 <sup>±0.6</sup>	69.8 <sup>±3.2</sup>	68.2 <sup>±0.5</sup>	72.0 <sup>±0.5</sup>	64.1 <sup>±0.2</sup>	70.7 <sup>±0.9</sup>	71.8 <sup>±1.1</sup>	69.9 <sup>±1.0</sup>
		20	68.5 <sup>±0.5</sup>	71.4 <sup>±1.3</sup>	64.5 <sup>±0.8</sup>	71.9 <sup>±0.9</sup>	72.7 <sup>±1.4</sup>	68.9 <sup>±2.7</sup>	66.5 <sup>±1.6</sup>	71.7 <sup>±0.5</sup>	65.2 <sup>±1.6</sup>	71.4 <sup>±0.5</sup>	72.4 <sup>±0.7</sup>	70.1 <sup>±2.0</sup>
		100	69.9 <sup>±0.6</sup>	73.9 <sup>±0.5</sup>	65.7 <sup>±0.9</sup>	71.7 <sup>±0.7</sup>	71.9 <sup>±0.9</sup>	76.0 <sup>±2.5</sup>	69.4 <sup>±0.8</sup>	73.7 <sup>±0.6</sup>	66.9 <sup>±1.2</sup>	71.1 <sup>±0.8</sup>	72.2 <sup>±0.6</sup>	73.8 <sup>±0.4</sup>
E2E-ABSA	Mono	73.4 <sup>±0.8</sup>	74.4 <sup>±0.6</sup>	69.9 <sup>±0.5</sup>	71.6 <sup>±1.0</sup>	72.4 <sup>±0.2</sup>	60.1 <sup>±1.7</sup>	73.5 <sup>±0.4</sup>	75.3 <sup>±0.6</sup>	69.8 <sup>±1.4</sup>	67.0 <sup>±0.4</sup>	72.2 <sup>±0.4</sup>	60.7 <sup>±1.1</sup>	
	Cross-lingual	0	57.3 <sup>±1.4</sup>	59.2 <sup>±0.5</sup>	57.8 <sup>±1.2</sup>	57.1 <sup>±0.9</sup>	56.4 <sup>±2.1</sup>	44.4 <sup>±1.4</sup>	62.4 <sup>±1.6</sup>	69.3 <sup>±1.0</sup>	61.1 <sup>±1.2</sup>	60.8 <sup>±0.3</sup>	63.7 <sup>±1.3</sup>	48.9 <sup>±1.4</sup>
		2	59.0 <sup>±1.1</sup>	61.1 <sup>±0.9</sup>	60.0 <sup>±1.3</sup>	59.0 <sup>±0.7</sup>	59.4 <sup>±0.8</sup>	49.2 <sup>±1.2</sup>	62.3 <sup>±0.9</sup>	70.5 <sup>±1.0</sup>	60.2 <sup>±0.3</sup>	61.7 <sup>±0.9</sup>	63.4 <sup>±1.0</sup>	51.4 <sup>±3.0</sup>
		5	58.9 <sup>±0.6</sup>	62.8 <sup>±2.7</sup>	59.3 <sup>±1.1</sup>	61.0 <sup>±0.4</sup>	59.6 <sup>±0.6</sup>	51.2 <sup>±1.9</sup>	63.0 <sup>±0.5</sup>	71.3 <sup>±0.7</sup>	60.4 <sup>±1.2</sup>	62.1 <sup>±1.1</sup>	63.7 <sup>±0.9</sup>	52.8 <sup>±1.8</sup>
		5	59.9 <sup>±1.1</sup>	63.2 <sup>±2.4</sup>	61.3 <sup>±1.1</sup>	60.5 <sup>±1.0</sup>	61.2 <sup>±0.7</sup>	54.8 <sup>±3.0</sup>	63.2 <sup>±0.3</sup>	71.0 <sup>±1.6</sup>	61.4 <sup>±0.6</sup>	61.1 <sup>±2.0</sup>	65.8 <sup>±0.8</sup>	53.2 <sup>±0.9</sup>
		10	60.7 <sup>±1.0</sup>	70.7 <sup>±1.3</sup>	60.9 <sup>±0.6</sup>	61.9 <sup>±1.3</sup>	60.0 <sup>±1.3</sup>	53.2 <sup>±1.7</sup>	63.6 <sup>±0.7</sup>	71.6 <sup>±0.9</sup>	61.4 <sup>±0.7</sup>	62.7 <sup>±1.5</sup>	65.8 <sup>±1.2</sup>	54.0 <sup>±1.6</sup>
		20	62.1 <sup>±1.2</sup>	71.2 <sup>±1.7</sup>	61.1 <sup>±0.7</sup>	62.1 <sup>±1.5</sup>	64.6 <sup>±1.7</sup>	52.5 <sup>±1.9</sup>	64.0 <sup>±1.6</sup>	72.1 <sup>±1.0</sup>	62.2 <sup>±1.0</sup>	63.2 <sup>±1.3</sup>	66.6 <sup>±1.0</sup>	56.2 <sup>±2.7</sup>
100	66.6 <sup>±0.4</sup>	72.2 <sup>±0.5</sup>	62.7 <sup>±1.0</sup>	65.8 <sup>±1.6</sup>	66.0 <sup>±1.6</sup>	57.7 <sup>±1.8</sup>	68.0 <sup>±0.6</sup>	73.1 <sup>±0.3</sup>	62.3 <sup>±0.4</sup>	64.9 <sup>±0.5</sup>	67.6 <sup>±0.8</sup>	58.4 <sup>±1.6</sup>		
ACTE	Mono	73.5 <sup>±0.8</sup>	70.4 <sup>±0.7</sup>	63.7 <sup>±0.8</sup>	68.8 <sup>±0.5</sup>	73.2 <sup>±0.5</sup>	59.1 <sup>±0.5</sup>	73.6 <sup>±0.5</sup>	69.9 <sup>±0.4</sup>	64.9 <sup>±0.5</sup>	62.9 <sup>±0.5</sup>	72.8 <sup>±1.0</sup>	60.1 <sup>±2.1</sup>	
	Cross-lingual	0	54.3 <sup>±1.6</sup>	52.5 <sup>±1.0</sup>	55.8 <sup>±0.7</sup>	52.3 <sup>±1.3</sup>	55.0 <sup>±2.7</sup>	41.4 <sup>±1.4</sup>	58.7 <sup>±1.0</sup>	62.8 <sup>±1.4</sup>	57.5 <sup>±0.3</sup>	54.1 <sup>±0.2</sup>	60.4 <sup>±0.9</sup>	49.0 <sup>±0.9</sup>
		2	55.2 <sup>±0.9</sup>	58.5 <sup>±0.8</sup>	56.2 <sup>±0.3</sup>	52.7 <sup>±1.1</sup>	57.1 <sup>±2.1</sup>	45.0 <sup>±2.1</sup>	58.8 <sup>±0.8</sup>	62.8 <sup>±0.2</sup>	55.8 <sup>±0.5</sup>	53.9 <sup>±0.8</sup>	61.5 <sup>±0.5</sup>	46.9 <sup>±1.1</sup>
		2	55.8 <sup>±1.2</sup>	57.7 <sup>±1.8</sup>	55.0 <sup>±0.7</sup>	52.1 <sup>±0.6</sup>	57.5 <sup>±1.1</sup>	46.4 <sup>±0.7</sup>	58.9 <sup>±1.0</sup>	62.8 <sup>±0.5</sup>	57.1 <sup>±0.9</sup>	53.4 <sup>±0.9</sup>	62.0 <sup>±0.4</sup>	47.1 <sup>±0.8</sup>
		5	57.0 <sup>±1.0</sup>	58.6 <sup>±0.4</sup>	56.3 <sup>±0.4</sup>	53.9 <sup>±0.9</sup>	58.1 <sup>±1.0</sup>	48.0 <sup>±3.4</sup>	59.7 <sup>±0.8</sup>	63.2 <sup>±1.0</sup>	58.0 <sup>±0.7</sup>	53.4 <sup>±1.6</sup>	62.8 <sup>±0.6</sup>	48.9 <sup>±1.9</sup>
		10	56.8 <sup>±1.3</sup>	63.3 <sup>±0.3</sup>	57.4 <sup>±1.6</sup>	56.1 <sup>±1.2</sup>	59.3 <sup>±1.3</sup>	49.7 <sup>±0.7</sup>	60.3 <sup>±1.3</sup>	63.9 <sup>±0.9</sup>	57.0 <sup>±0.8</sup>	55.7 <sup>±1.7</sup>	63.2 <sup>±0.8</sup>	50.7 <sup>±1.3</sup>
		20	57.4 <sup>±1.6</sup>	64.6 <sup>±0.5</sup>	57.8 <sup>±0.8</sup>	55.2 <sup>±1.2</sup>	63.5 <sup>±0.8</sup>	49.8 <sup>±1.2</sup>	61.0 <sup>±1.1</sup>	64.5 <sup>±1.0</sup>	57.3 <sup>±0.7</sup>	55.5 <sup>±0.5</sup>	64.2 <sup>±1.2</sup>	51.1 <sup>±1.7</sup>
100	63.6 <sup>±0.7</sup>	65.9 <sup>±0.6</sup>	59.8 <sup>±0.4</sup>	60.1 <sup>±1.0</sup>	66.0 <sup>±0.7</sup>	53.6 <sup>±1.2</sup>	64.4 <sup>±1.1</sup>	66.4 <sup>±1.0</sup>	59.7 <sup>±0.9</sup>	58.5 <sup>±0.5</sup>	66.2 <sup>±0.8</sup>	55.2 <sup>±1.6</sup>		
TASD	Mono	66.9 <sup>±0.3</sup>	65.8 <sup>±0.4</sup>	59.0 <sup>±0.6</sup>	62.9 <sup>±1.4</sup>	67.0 <sup>±0.9</sup>	54.1 <sup>±3.0</sup>	67.1 <sup>±1.3</sup>	66.2 <sup>±0.5</sup>	58.9 <sup>±1.1</sup>	57.6 <sup>±0.5</sup>	66.4 <sup>±0.4</sup>	53.9 <sup>±1.5</sup>	
	Cross-lingual	0	50.2 <sup>±0.9</sup>	48.3 <sup>±0.5</sup>	50.4 <sup>±1.4</sup>	47.7 <sup>±1.1</sup>	48.6 <sup>±2.0</sup>	39.1 <sup>±3.6</sup>	53.3 <sup>±1.5</sup>	57.6 <sup>±0.6</sup>	50.4 <sup>±0.8</sup>	50.4 <sup>±1.3</sup>	54.9 <sup>±2.0</sup>	43.8 <sup>±0.8</sup>
		1	49.9 <sup>±0.4</sup>	52.5 <sup>±1.2</sup>	50.0 <sup>±1.2</sup>	48.1 <sup>±0.8</sup>	52.3 <sup>±1.6</sup>	42.0 <sup>±1.8</sup>	53.9 <sup>±0.9</sup>	58.0 <sup>±0.5</sup>	50.4 <sup>±1.0</sup>	49.8 <sup>±0.9</sup>	55.5 <sup>±0.5</sup>	43.9 <sup>±1.6</sup>
		2	50.7 <sup>±1.5</sup>	51.3 <sup>±1.5</sup>	49.6 <sup>±1.2</sup>	49.1 <sup>±0.6</sup>	53.1 <sup>±1.5</sup>	43.7 <sup>±1.9</sup>	53.9 <sup>±0.9</sup>	58.0 <sup>±0.6</sup>	50.9 <sup>±1.4</sup>	50.2 <sup>±0.3</sup>	55.5 <sup>±0.7</sup>	44.1 <sup>±1.7</sup>
		5	52.4 <sup>±1.7</sup>	52.4 <sup>±2.1</sup>	50.2 <sup>±1.2</sup>	49.7 <sup>±1.3</sup>	52.3 <sup>±1.4</sup>	45.1 <sup>±1.4</sup>	53.2 <sup>±1.1</sup>	58.4 <sup>±0.6</sup>	51.2 <sup>±1.6</sup>	49.0 <sup>±1.3</sup>	57.5 <sup>±0.4</sup>	46.5 <sup>±1.0</sup>
		10	52.5 <sup>±1.3</sup>	58.2 <sup>±1.0</sup>	49.8 <sup>±0.9</sup>	51.2 <sup>±1.5</sup>	53.6 <sup>±1.5</sup>	45.3 <sup>±2.7</sup>	54.1 <sup>±0.6</sup>	58.8 <sup>±0.6</sup>	51.0 <sup>±0.8</sup>	50.4 <sup>±1.2</sup>	57.5 <sup>±1.1</sup>	46.3 <sup>±1.5</sup>
		20	51.6 <sup>±0.6</sup>	58.9 <sup>±0.8</sup>	50.6 <sup>±1.3</sup>	52.5 <sup>±0.8</sup>	57.8 <sup>±1.0</sup>	44.8 <sup>±2.1</sup>	54.1 <sup>±1.3</sup>	60.2 <sup>±0.7</sup>	52.5 <sup>±1.4</sup>	52.3 <sup>±0.8</sup>	58.0 <sup>±1.3</sup>	47.4 <sup>±2.2</sup>
100	57.5 <sup>±0.4</sup>	61.5 <sup>±0.8</sup>	52.4 <sup>±1.1</sup>	54.1 <sup>±0.7</sup>	58.9 <sup>±0.9</sup>	49.7 <sup>±1.4</sup>	57.1 <sup>±0.7</sup>	61.7 <sup>±0.7</sup>	52.7 <sup>±1.6</sup>	54.3 <sup>±0.8</sup>	58.8 <sup>±1.4</sup>	50.1 <sup>±0.6</sup>		

Overall, we observe a clear trend of performance improvement with an increasing number of few-shot examples. This improvement is especially noticeable when constrained decoding is not used. For instance, in the TASD task with Spanish as the target language, adding a single few-shot example improves performance by approximately 4% over the zero-shot setting without constrained decoding. Increasing the number of examples to ten provides an additional 6% gain – a 10% improvement over the zero-shot baseline. Interestingly, as the number of few-shot examples increases, the relative benefit of constrained decoding decreases. With ten few-shot examples, constrained decoding usually provides no significant ad-

vantage. An exception is Czech, where constrained decoding consistently improves results, even with 100 few-shot examples – possibly due to specific morphological or syntactic characteristics of the language.

The ACSA task, which had not previously been evaluated with constrained decoding, shows no consistent improvement from its use. This aligns with expectations, as constrained decoding primarily addresses challenges related to aspect term prediction [15], which is not a component of ACSA.

Across tasks, the most substantial gains compared to zero-shot performance generally occur when moving to ten few-shot examples. Improvements with fewer examples (e.g. one or five) tend to be modest or not statistically significant. Increasing the count from ten to twenty examples typically yields to only a small or no further gains. A more substantial leap in performance is usually seen only when scaling up to 100 examples, suggesting a non-linear benefit from additional supervision. With 100 few-shot examples, performance is typically between 2–10% lower than monolingual results, depending on the task, while with ten few-shot examples, performance is usually about 3% lower than with 100 examples.

Table 4 shows analogous results using the mBART model. While the overall patterns mirror those observed with mT5, a notable difference emerges: constrained decoding has significantly less impact on mBART. It generally leads to improvements only in a subset of languages under zero-shot settings. This suggests that mBART may be inherently more robust to output structure violations or benefits less from structural constraints. However, the overall results are better with mT5 than with mBART.

In summary, even a small number of high-quality few-shot examples – particularly ten – can yield substantial gains over zero-shot performance, often surpassing zero-shot constrained decoding results. Given that collecting ten labelled instances per target language is a manageable effort, few-shot learning presents a highly practical and efficient approach for cross-lingual ABSA. Moreover, the diminishing returns beyond ten examples – particularly when weighed against the increased time and cost of data labelling – underscore the efficiency of small-scale supervision and offer promising implications for low-resource or domain-specific adaptation scenarios.

### 5.1 TASD with Different Few-Shot Examples

We investigate the impact of increasing the number of target-language examples in the training data for the TASD task, with results shown in Figure 2. As expected, adding more examples in the target language generally improves performance. For most languages, there is a clear upward trend, with models often approaching or even surpassing monolingual baselines. However, the gains tend to plateau around 1,000 examples, highlighting a practical ceiling given the cost of obtaining high-quality annotations in multiple languages.

As discussed previously, even a small addition of ten examples can yield substantial improvements over the zero-shot setting. Interestingly, this benefit is often more pronounced when constrained decoding is not used. This suggests

Table 4: Cross-lingual micro F1-scores on four tasks in six target languages using mBART, compared to monolingual (mono) performance, with varying few-shot examples (FS). **Bold** marks better constrained decoding (CD) results with non-overlapping 95% confidence intervals. Asterisk (\*) results are from [15].

Task Setup FS		Without constrained decoding						With constrained decoding						
		Cs	Es	Fr	Nl	Ru	Tr	Cs	Es	Fr	Nl	Ru	Tr	
ACSA	Mono	72.6 <sup>±1.3</sup>	73.2 <sup>±1.4</sup>	65.0 <sup>±0.9</sup>	70.2 <sup>±1.9</sup>	73.3 <sup>±1.2</sup>	66.1 <sup>±4.0</sup>	71.5 <sup>±0.8</sup>	73.3 <sup>±0.8</sup>	63.7 <sup>±1.0</sup>	68.3 <sup>±2.9</sup>	72.8 <sup>±1.0</sup>	64.4 <sup>±5.1</sup>	
	Cross-lingual	0	55.2 <sup>±1.4</sup>	61.7 <sup>±2.5</sup>	53.7 <sup>±2.4</sup>	58.4 <sup>±2.1</sup>	65.6 <sup>±1.9</sup>	50.7 <sup>±2.6</sup>	53.8 <sup>±4.0</sup>	63.0 <sup>±2.3</sup>	53.0 <sup>±1.8</sup>	59.4 <sup>±2.0</sup>	66.4 <sup>±1.7</sup>	49.3 <sup>±2.7</sup>
		1	56.6 <sup>±1.2</sup>	64.1 <sup>±1.0</sup>	54.9 <sup>±1.0</sup>	60.1 <sup>±0.9</sup>	68.2 <sup>±0.9</sup>	54.4 <sup>±1.1</sup>	57.4 <sup>±1.1</sup>	64.2 <sup>±0.5</sup>	54.0 <sup>±0.8</sup>	61.1 <sup>±1.8</sup>	68.5 <sup>±1.1</sup>	54.8 <sup>±0.7</sup>
		2	56.7 <sup>±1.4</sup>	64.4 <sup>±1.1</sup>	55.7 <sup>±1.1</sup>	60.9 <sup>±0.8</sup>	67.2 <sup>±1.3</sup>	52.1 <sup>±1.7</sup>	56.3 <sup>±1.6</sup>	65.2 <sup>±1.3</sup>	55.0 <sup>±1.5</sup>	60.8 <sup>±0.9</sup>	68.1 <sup>±0.8</sup>	53.2 <sup>±2.0</sup>
		5	56.4 <sup>±0.9</sup>	63.9 <sup>±1.8</sup>	56.2 <sup>±1.2</sup>	61.3 <sup>±1.3</sup>	67.1 <sup>±0.6</sup>	51.6 <sup>±4.8</sup>	56.2 <sup>±0.7</sup>	65.0 <sup>±1.4</sup>	55.9 <sup>±2.0</sup>	61.5 <sup>±1.1</sup>	67.6 <sup>±0.7</sup>	53.7 <sup>±3.5</sup>
		10	57.9 <sup>±2.5</sup>	62.5 <sup>±3.1</sup>	55.7 <sup>±1.3</sup>	58.4 <sup>±2.2</sup>	66.5 <sup>±1.8</sup>	54.4 <sup>±1.9</sup>	55.5 <sup>±1.5</sup>	62.8 <sup>±2.0</sup>	54.0 <sup>±3.3</sup>	58.3 <sup>±3.9</sup>	67.6 <sup>±0.5</sup>	53.4 <sup>±5.8</sup>
		20	57.0 <sup>±3.4</sup>	64.0 <sup>±2.3</sup>	57.2 <sup>±2.4</sup>	62.8 <sup>±0.7</sup>	65.9 <sup>±1.5</sup>	55.5 <sup>±3.2</sup>	56.0 <sup>±4.1</sup>	61.5 <sup>±1.8</sup>	56.9 <sup>±2.1</sup>	60.3 <sup>±3.7</sup>	66.0 <sup>±0.7</sup>	55.9 <sup>±4.3</sup>
100	63.5 <sup>±1.5</sup>	67.6 <sup>±2.1</sup>	59.4 <sup>±2.2</sup>	64.5 <sup>±1.9</sup>	67.1 <sup>±3.1</sup>	61.9 <sup>±4.6</sup>	62.4 <sup>±2.0</sup>	66.5 <sup>±0.9</sup>	59.3 <sup>±1.5</sup>	65.0 <sup>±1.6</sup>	65.5 <sup>±2.3</sup>	62.6 <sup>±4.7</sup>		
EPE-ABSA	Mono	69.1 <sup>±0.3</sup>	73.0 <sup>±0.5</sup>	66.4 <sup>±1.1</sup>	68.9 <sup>±1.2</sup>	68.7 <sup>±1.6</sup>	56.0 <sup>±2.7</sup>	68.8 <sup>±0.8</sup>	71.9 <sup>±1.3</sup>	64.0 <sup>±1.7</sup>	61.6 <sup>±1.0</sup>	66.2 <sup>±1.1</sup>	54.4 <sup>±2.3</sup>	
	Cross-lingual	0	51.5 <sup>±3.3</sup>	61.1 <sup>±2.6</sup>	49.4 <sup>±3.8</sup>	51.6 <sup>±2.7</sup>	57.1 <sup>±1.4</sup>	31.6 <sup>±3.9</sup>	48.8 <sup>±2.7</sup>	61.7 <sup>±2.7</sup>	49.2 <sup>±4.1</sup>	50.1 <sup>±3.5</sup>	57.8 <sup>±1.8</sup>	30.3 <sup>±3.0</sup>
		1	49.2 <sup>±1.2</sup>	56.6 <sup>±0.6</sup>	49.9 <sup>±0.5</sup>	47.3 <sup>±0.7</sup>	55.8 <sup>±1.2</sup>	31.8 <sup>±1.5</sup>	49.7 <sup>±0.8</sup>	57.1 <sup>±1.6</sup>	48.9 <sup>±1.1</sup>	46.1 <sup>±0.7</sup>	56.9 <sup>±1.2</sup>	32.6 <sup>±2.4</sup>
		2	50.1 <sup>±2.0</sup>	55.7 <sup>±1.1</sup>	50.1 <sup>±1.6</sup>	47.7 <sup>±1.1</sup>	57.5 <sup>±1.6</sup>	32.6 <sup>±2.6</sup>	49.4 <sup>±1.6</sup>	56.5 <sup>±0.6</sup>	48.6 <sup>±0.5</sup>	47.9 <sup>±1.8</sup>	56.7 <sup>±1.2</sup>	32.7 <sup>±1.6</sup>
		5	51.8 <sup>±0.7</sup>	57.0 <sup>±1.1</sup>	51.3 <sup>±1.0</sup>	49.4 <sup>±0.6</sup>	57.5 <sup>±1.2</sup>	30.7 <sup>±0.8</sup>	50.6 <sup>±1.4</sup>	56.4 <sup>±0.8</sup>	50.2 <sup>±1.5</sup>	47.9 <sup>±1.4</sup>	56.2 <sup>±0.4</sup>	31.7 <sup>±2.3</sup>
		10	51.0 <sup>±3.4</sup>	64.1 <sup>±2.2</sup>	53.4 <sup>±2.0</sup>	53.4 <sup>±2.9</sup>	58.8 <sup>±1.7</sup>	37.7 <sup>±3.2</sup>	52.0 <sup>±3.4</sup>	64.8 <sup>±2.6</sup>	51.6 <sup>±1.1</sup>	51.4 <sup>±2.9</sup>	57.9 <sup>±1.8</sup>	36.3 <sup>±2.0</sup>
		20	53.6 <sup>±2.9</sup>	63.2 <sup>±2.8</sup>	54.3 <sup>±1.8</sup>	53.4 <sup>±2.0</sup>	59.5 <sup>±1.6</sup>	40.9 <sup>±3.9</sup>	52.8 <sup>±3.4</sup>	65.5 <sup>±1.3</sup>	51.7 <sup>±1.1</sup>	53.6 <sup>±1.2</sup>	58.9 <sup>±2.4</sup>	36.4 <sup>±1.9</sup>
100	60.6 <sup>±1.3</sup>	67.6 <sup>±1.4</sup>	57.2 <sup>±1.2</sup>	57.9 <sup>±1.7</sup>	62.3 <sup>±1.9</sup>	47.2 <sup>±3.1</sup>	59.7 <sup>±1.7</sup>	67.5 <sup>±1.1</sup>	55.3 <sup>±0.8</sup>	56.1 <sup>±1.6</sup>	60.5 <sup>±0.8</sup>	46.6 <sup>±3.7</sup>		
ACTE	Mono	70.1 <sup>±0.8</sup>	66.4 <sup>±1.6</sup>	61.1 <sup>±1.6</sup>	64.1 <sup>±1.2</sup>	70.9 <sup>±0.6</sup>	56.8 <sup>±2.2</sup>	68.4 <sup>±0.9</sup>	66.8 <sup>±1.5</sup>	58.2 <sup>±1.2</sup>	58.0 <sup>±1.2</sup>	67.4 <sup>±0.3</sup>	55.3 <sup>±1.5</sup>	
	Cross-lingual	0	48.6 <sup>±2.7</sup>	52.5 <sup>±1.4</sup>	49.3 <sup>±1.5</sup>	44.5 <sup>±1.4</sup>	53.8 <sup>±1.5</sup>	31.1 <sup>±2.1</sup>	45.8 <sup>±4.5</sup>	54.8 <sup>±0.4</sup>	49.2 <sup>±0.6</sup>	46.9 <sup>±0.9</sup>	55.9 <sup>±0.2</sup>	34.7 <sup>±1.1</sup>
		1	52.7 <sup>±1.0</sup>	63.0 <sup>±1.0</sup>	52.3 <sup>±2.1</sup>	52.2 <sup>±0.9</sup>	58.0 <sup>±0.7</sup>	33.6 <sup>±2.2</sup>	52.1 <sup>±1.8</sup>	64.5 <sup>±1.1</sup>	51.8 <sup>±1.3</sup>	52.6 <sup>±0.9</sup>	58.8 <sup>±1.7</sup>	34.5 <sup>±1.6</sup>
		2	52.7 <sup>±1.2</sup>	63.3 <sup>±0.7</sup>	51.6 <sup>±0.9</sup>	52.1 <sup>±2.2</sup>	59.2 <sup>±0.8</sup>	35.2 <sup>±1.0</sup>	51.8 <sup>±2.0</sup>	64.7 <sup>±0.5</sup>	50.8 <sup>±1.0</sup>	52.1 <sup>±1.4</sup>	59.0 <sup>±0.8</sup>	34.7 <sup>±1.2</sup>
		5	50.9 <sup>±0.9</sup>	63.6 <sup>±0.6</sup>	54.1 <sup>±1.0</sup>	52.5 <sup>±0.5</sup>	59.6 <sup>±0.7</sup>	36.4 <sup>±1.0</sup>	51.8 <sup>±2.1</sup>	64.9 <sup>±0.7</sup>	53.4 <sup>±1.5</sup>	51.9 <sup>±0.5</sup>	59.1 <sup>±0.5</sup>	35.5 <sup>±0.9</sup>
		10	52.5 <sup>±0.9</sup>	57.3 <sup>±1.1</sup>	50.6 <sup>±1.9</sup>	47.4 <sup>±2.2</sup>	56.0 <sup>±2.3</sup>	35.1 <sup>±2.9</sup>	47.5 <sup>±0.8</sup>	56.4 <sup>±1.5</sup>	48.5 <sup>±0.5</sup>	45.9 <sup>±1.3</sup>	55.7 <sup>±1.9</sup>	33.6 <sup>±5.1</sup>
		20	50.8 <sup>±1.0</sup>	59.4 <sup>±0.6</sup>	50.9 <sup>±1.5</sup>	49.8 <sup>±2.1</sup>	59.7 <sup>±0.9</sup>	41.1 <sup>±2.9</sup>	51.6 <sup>±2.8</sup>	57.8 <sup>±1.6</sup>	47.8 <sup>±1.6</sup>	48.8 <sup>±2.8</sup>	59.1 <sup>±1.6</sup>	39.6 <sup>±4.4</sup>
100	59.2 <sup>±1.6</sup>	60.4 <sup>±2.5</sup>	53.3 <sup>±2.2</sup>	55.0 <sup>±2.4</sup>	62.1 <sup>±1.0</sup>	46.6 <sup>±4.9</sup>	57.1 <sup>±2.2</sup>	61.0 <sup>±1.8</sup>	51.1 <sup>±1.1</sup>	52.5 <sup>±2.1</sup>	60.3 <sup>±1.9</sup>	44.8 <sup>±3.9</sup>		
TASD	Mono	62.6 <sup>±0.7</sup>	62.9 <sup>±1.2</sup>	54.8 <sup>±0.9</sup>	57.6 <sup>±0.9</sup>	62.6 <sup>±0.7</sup>	49.3 <sup>±3.1</sup>	61.9 <sup>±1.6</sup>	61.5 <sup>±1.4</sup>	52.4 <sup>±0.6</sup>	52.1 <sup>±1.0</sup>	60.1 <sup>±1.9</sup>	47.6 <sup>±2.7</sup>	
	Cross-lingual	0	40.4 <sup>±3.0</sup>	47.6 <sup>±1.9</sup>	39.6 <sup>±0.8</sup>	39.1 <sup>±0.9</sup>	48.5 <sup>±1.1</sup>	23.5 <sup>±2.6</sup>	39.3 <sup>±1.0</sup>	51.1 <sup>±1.2</sup>	39.9 <sup>±0.6</sup>	38.9 <sup>±0.9</sup>	50.5 <sup>±0.7</sup>	27.3 <sup>±1.1</sup>
		1	42.8 <sup>±1.2</sup>	50.4 <sup>±1.8</sup>	41.3 <sup>±1.2</sup>	40.9 <sup>±0.8</sup>	51.2 <sup>±0.4</sup>	26.8 <sup>±1.2</sup>	41.1 <sup>±1.0</sup>	51.5 <sup>±0.9</sup>	40.5 <sup>±0.9</sup>	39.2 <sup>±0.7</sup>	50.8 <sup>±0.9</sup>	27.2 <sup>±1.4</sup>
		2	42.4 <sup>±1.2</sup>	50.5 <sup>±1.0</sup>	42.3 <sup>±0.9</sup>	40.6 <sup>±1.1</sup>	50.8 <sup>±0.7</sup>	28.9 <sup>±3.3</sup>	42.0 <sup>±0.9</sup>	51.3 <sup>±1.3</sup>	39.7 <sup>±1.1</sup>	40.1 <sup>±2.1</sup>	50.6 <sup>±1.2</sup>	29.1 <sup>±2.7</sup>
		5	42.8 <sup>±1.5</sup>	52.3 <sup>±1.5</sup>	42.1 <sup>±1.2</sup>	41.6 <sup>±1.1</sup>	50.0 <sup>±0.2</sup>	27.3 <sup>±3.2</sup>	42.2 <sup>±1.9</sup>	51.3 <sup>±1.0</sup>	41.0 <sup>±0.8</sup>	41.8 <sup>±1.1</sup>	51.5 <sup>±0.9</sup>	26.8 <sup>±1.4</sup>
		10	43.3 <sup>±2.3</sup>	51.8 <sup>±1.0</sup>	40.9 <sup>±3.8</sup>	42.2 <sup>±3.1</sup>	49.9 <sup>±1.0</sup>	29.8 <sup>±4.4</sup>	42.4 <sup>±2.2</sup>	51.7 <sup>±3.3</sup>	40.1 <sup>±1.7</sup>	39.7 <sup>±3.5</sup>	50.0 <sup>±0.7</sup>	28.3 <sup>±1.8</sup>
		20	43.5 <sup>±1.6</sup>	53.8 <sup>±2.2</sup>	42.0 <sup>±0.6</sup>	42.1 <sup>±2.4</sup>	51.6 <sup>±2.2</sup>	32.2 <sup>±0.6</sup>	42.0 <sup>±3.1</sup>	51.4 <sup>±1.8</sup>	40.3 <sup>±2.5</sup>	41.3 <sup>±2.1</sup>	51.5 <sup>±2.0</sup>	29.9 <sup>±2.5</sup>
100	51.3 <sup>±0.9</sup>	56.1 <sup>±2.7</sup>	44.9 <sup>±1.2</sup>	48.2 <sup>±2.4</sup>	54.3 <sup>±0.7</sup>	40.8 <sup>±2.9</sup>	49.7 <sup>±1.0</sup>	55.8 <sup>±0.3</sup>	42.4 <sup>±1.8</sup>	46.8 <sup>±1.9</sup>	52.0 <sup>±1.2</sup>	39.8 <sup>±1.8</sup>		

that a few target-language examples may help compensate for typical generation errors, such as producing aspect terms in the wrong language – an issue that constrained decoding was designed to mitigate.

With more target-language data, the relative advantage of constrained decoding gradually diminishes. Below 100 examples, constrained decoding can still offer some improvements, but its impact becomes marginal as the number increases. A notable exception is Dutch, where performance with constrained decoding consistently lags behind as more examples are added. In contrast, Dutch models without constrained decoding show marked improvements once the data reaches 200 examples or more.

## 5.2 Error Analysis

We conduct an error analysis to identify the most challenging sentiment prediction elements. We manually examine 100 random test samples from the best-performing mT5 runs – with and without constrained decoding – focusing on Czech and Spanish for TASD with up to 100 few-shot examples.

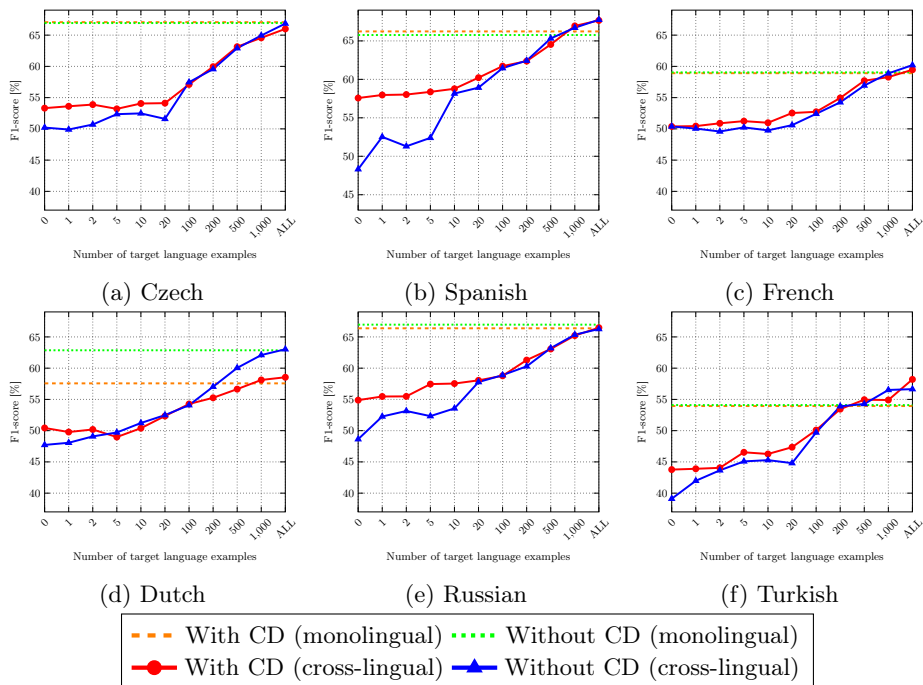


Fig. 2: Effect of adding target language examples on cross-lingual TASD performance with mT5, with and without constrained decoding (CD), compared to monolingual models.

The most frequent errors involve aspect term prediction. As noted in [15], the model sometimes outputs aspect terms in the source language instead of the target. This issue, along with typos correction (e.g. “*sevrice*” instead of “*service*”) and hallucinated words, is reduced by constrained decoding and few-shot examples. We also observe incomplete, irrelevant, or missing aspect terms. Error rates tend to decrease with more target-language few-shot examples, aligning with overall performance trends.

Aspect category errors are less common. Rare categories like “*drinks prices*” are often missed, and similar ones such as “*restaurant general*” and “*restaurant miscellaneous*” are frequently confused. Some categories, like “*food general*”, appear only in one language, which hinders cross-lingual transfer.

Sentiment polarity errors are the least frequent and mainly involve misclassifying the “*neutral*” class, likely due to label imbalance, since “*neutral*” is the least frequent class across all datasets.

## 6 Conclusion

This paper investigates the effect of incorporating few-shot target language examples into training data for cross-lingual aspect-based sentiment analysis using sequence-to-sequence models. Across four ABSA tasks, six target languages, and two multilingual models, we show that even a small number of target language examples – particularly ten – can lead to significant performance improvements, often rendering techniques like constrained decoding unnecessary. With larger few-shot sets, performance can exceed monolingual baselines, highlighting the strong potential of minimal in-language supervision. These findings offer a practical and cost-effective alternative to zero-shot cross-lingual approaches, especially valuable in low-resource and domain-specific scenarios where obtaining a handful of high-quality annotations is feasible.

## Acknowledgements

This work has been partly supported by the project R&D of Technologies for Advanced Digitalization in the Pilsen Metropolitan Area (DigiTech) No. CZ.02.01.01/00/23\_021/0008436. Computational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic.

## References

1. Barnes, J., Lambert, P., Badia, T.: Exploring distributional representations and machine translation for aspect-based cross-lingual sentiment classification. In: Matsumoto, Y., Prasad, R. (eds.) Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. pp. 1613–1623. The COLING 2016 Organizing Committee, Osaka, Japan (Dec 2016), <https://aclanthology.org/C16-1152>
2. Cao, N.D., Izacard, G., Riedel, S., Petroni, F.: Autoregressive entity retrieval. In: 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021. OpenReview.net (2021), <https://openreview.net/forum?id=5k8F6UU39V>
3. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V.: Unsupervised cross-lingual representation learning at scale. In: Jurafsky, D., Chai, J., Schluter, N., Tetreault, J. (eds.) Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 8440–8451. Association for Computational Linguistics, Online (Jul 2020). <https://doi.org/10.18653/v1/2020.acl-main.747>, <https://aclanthology.org/2020.acl-main.747>
4. Jebbara, S., Cimiano, P.: Zero-shot cross-lingual opinion target extraction. In: Burstein, J., Doran, C., Solorio, T. (eds.) Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). pp. 2486–2495. Association for Computational Linguistics, Minneapolis, Minnesota (Jun 2019). <https://doi.org/10.18653/v1/N19-1257>, <https://aclanthology.org/N19-1257>

5. Lambert, P.: Aspect-level cross-lingual sentiment classification with constrained SMT. In: Zong, C., Strube, M. (eds.) Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). pp. 781–787. Association for Computational Linguistics, Beijing, China (Jul 2015). <https://doi.org/10.3115/v1/P15-2128>, <https://aclanthology.org/P15-2128>
6. Li, X., Bing, L., Zhang, W., Li, Z., Lam, W.: Unsupervised cross-lingual adaptation for sequence tagging and beyond. arXiv preprint arXiv:2010.12405 (2020)
7. Lin, N., Fu, Y., Lin, X., Zhou, D., Yang, A., Jiang, S.: Cl-xabsa: Contrastive learning for cross-lingual aspect-based sentiment analysis. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **31**, 2935–2946 (2023). <https://doi.org/10.1109/TASLP.2023.3297964>
8. Lin, N., Zeng, M., Liao, X., Liu, W., Yang, A., Zhou, D.: Addressing class-imbalance challenges in cross-lingual aspect-based sentiment analysis: Dynamic weighted loss and anti-decoupling. *Expert Systems with Applications* **257**, 125059 (2024). <https://doi.org/10.1016/j.eswa.2024.125059>, <https://www.sciencedirect.com/science/article/pii/S0957417424019262>
9. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6–9, 2019. OpenReview.net (2019), <https://openreview.net/forum?id=Bkg6RiCqY7>
10. Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S.M., Eryigit, G.: SemEval-2016 task 5: Aspect based sentiment analysis. In: Bethard, S., Carpuat, M., Cer, D., Jurgens, D., Nakov, P., Zesch, T. (eds.) Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). pp. 19–30. Association for Computational Linguistics, San Diego, California (Jun 2016). <https://doi.org/10.18653/v1/S16-1002>, <https://aclanthology.org/S16-1002>
11. Schmitt, M., Steinheber, S., Schreiber, K., Roth, B.: Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks. In: Riloff, E., Chiang, D., Hockenmaier, J., Tsujii, J. (eds.) Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. pp. 1109–1114. Association for Computational Linguistics, Brussels, Belgium (Oct–Nov 2018). <https://doi.org/10.18653/v1/D18-1139>, <https://aclanthology.org/D18-1139/>
12. Shazeer, N., Stern, M.: Adafactor: Adaptive learning rates with sublinear memory cost. In: Dy, J.G., Krause, A. (eds.) Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholm, Sweden, July 10–15, 2018. Proceedings of Machine Learning Research, vol. 80, pp. 4603–4611. PMLR (2018), <http://proceedings.mlr.press/v80/shazeer18a.html>
13. Šmíd, J., Král, P.: Cross-lingual aspect-based sentiment analysis: A survey on tasks, approaches, and challenges. *Information Fusion* **120**, 103073 (2025). <https://doi.org/10.1016/j.inffus.2025.103073>, <https://www.sciencedirect.com/science/article/pii/S1566253525001460>
14. Šmíd, J., Priban, P., Kral, P.: LLaMA-based models for aspect-based sentiment analysis. In: De Clercq, O., Barriere, V., Barnes, J., Klinger, R., Sedoc, J., Tafreshi, S. (eds.) Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis. pp. 63–70. Association for Computational Linguistics, Bangkok, Thailand (Aug 2024).

- <https://doi.org/10.18653/v1/2024.wassa-1.6>, <https://aclanthology.org/2024.wassa-1.6/>
15. Šmíd, J., Přibáň, P., Král, P.: Advancing cross-lingual aspect-based sentiment analysis with llms and constrained decoding for sequence-to-sequence models. In: Proceedings of the 17th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART. pp. 757–766. INSTICC, SciTePress (2025). <https://doi.org/10.5220/0013349400003890>
  16. Šmíd, J., Přibáň, P., Prazak, O., Kral, P.: Czech dataset for complex aspect-based sentiment analysis tasks. In: Calzolari, N., Kan, M.Y., Hoste, V., Lenci, A., Sakti, S., Xue, N. (eds.) Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). pp. 4299–4310. ELRA and ICCL, Torino, Italia (May 2024), <https://aclanthology.org/2024.lrec-main.384/>
  17. Tang, Y., Tran, C., Li, X., Chen, P.J., Goyal, N., Chaudhary, V., Gu, J., Fan, A.: Multilingual translation with extensible multilingual pretraining and finetuning. arXiv preprint arXiv:2008.00401 (2020)
  18. Wan, H., Yang, Y., Du, J., Liu, Y., Qi, K., Pan, J.Z.: Target-aspect-sentiment joint detection for aspect-based sentiment analysis. Proceedings of the AAAI Conference on Artificial Intelligence **34**(05), 9122–9129 (Apr 2020). <https://doi.org/10.1609/aaai.v34i05.6447>, <https://ojs.aaai.org/index.php/AAAI/article/view/6447>
  19. Wang, F., Lan, M., Wang, W.: Towards a one-stop solution to both aspect extraction and sentiment analysis tasks with neural multi-task learning. In: 2018 International joint conference on neural networks (IJCNN). pp. 1–8. IEEE (2018)
  20. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., Rush, A.: Transformers: State-of-the-art natural language processing. In: Liu, Q., Schlangen, D. (eds.) Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. pp. 38–45. Association for Computational Linguistics, Online (Oct 2020). <https://doi.org/10.18653/v1/2020.emnlp-demos.6>, <https://aclanthology.org/2020.emnlp-demos.6/>
  21. Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., Raffel, C.: mT5: A massively multilingual pre-trained text-to-text transformer. In: Toutanova, K., Rumshisky, A., Zettlemoyer, L., Hakkani-Tur, D., Beltagy, I., Bethard, S., Cotterell, R., Chakraborty, T., Zhou, Y. (eds.) Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 483–498. Association for Computational Linguistics, Online (Jun 2021). <https://doi.org/10.18653/v1/2021.naacl-main.41>, <https://aclanthology.org/2021.naacl-main.41>
  22. Zhang, W., He, R., Peng, H., Bing, L., Lam, W.: Cross-lingual aspect-based sentiment analysis with aspect term code-switching. In: Moens, M.F., Huang, X., Specia, L., Yih, S.W.t. (eds.) Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. pp. 9220–9230. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic (Nov 2021). <https://doi.org/10.18653/v1/2021.emnlp-main.727>, <https://aclanthology.org/2021.emnlp-main.727/>