

Cross-lingual Aspect-Based Sentiment Analysis: A Survey on Tasks, Approaches, and Challenges^{*}

Jakub Šmíd^{a,b,*}, Pavel Král^{a,b}

^aUniversity of West Bohemia in Pilsen, Faculty of Applied Sciences, Department of Computer Science and Engineering, Univerzitní 8, Pilsen 301 00, Czech Republic

^bNTIS – New Technologies for the Information Society, Univerzitní 8, Pilsen 301 00, Czech Republic

ARTICLE INFO

Keywords:

Cross-lingual aspect-based sentiment analysis
Aspect-based sentiment analysis
Sentiment analysis
Opinion mining
Cross-lingual transfer
Machine learning
Pre-trained language models

ABSTRACT

Aspect-based sentiment analysis (ABSA) is a fine-grained sentiment analysis task that focuses on understanding opinions at the aspect level, including sentiment towards specific aspect terms, categories, and opinions. While ABSA research has seen significant progress, much of the focus has been on monolingual settings. Cross-lingual ABSA, which aims to transfer knowledge from resource-rich languages (such as English) to low-resource languages, remains an under-explored area, with no systematic review of the field. This paper aims to fill that gap by providing a comprehensive survey of cross-lingual ABSA. We summarize key ABSA tasks, including aspect term extraction, aspect sentiment classification, and compound tasks involving multiple sentiment elements. Additionally, we review the datasets, modelling paradigms, and cross-lingual transfer methods used to solve these tasks. We also examine how existing work in monolingual and multilingual ABSA, as well as ABSA with LLMs, contributes to the development of cross-lingual ABSA. Finally, we highlight the main challenges and suggest directions for future research to advance cross-lingual ABSA systems.

1. Introduction

Aspect-based sentiment analysis (ABSA), a natural language processing (NLP) task, has become essential for analysing opinions expressed in online content, especially in environments where customer feedback and reviews shape decisions. Businesses, for example, rely heavily on customer sentiment to refine their products, improve services, and develop marketing strategies. However, given the sheer volume of unstructured online text, manually processing these opinions is infeasible. This growing demand for scalable solutions has driven the need for automated systems capable of extracting and interpreting sentiment from user-generated content, leading to the rise of sentiment analysis and opinion mining [1].

Traditional sentiment analysis often focuses on broader evaluation of entire documents or sentences. These approaches assume that the text conveys a single sentiment towards a single topic, which oversimplifies the complexity of real-world opinions. In reality, different aspects of a product or service may evoke distinct sentiments within a single review. For instance, a customer might express satisfaction with a product's design but disappointment with its price. This gap led to the development of aspect-based sentiment analysis, which focuses on extracting sentiment at a more granular level. Rather than analysing the sentiment of the entire text, ABSA identifies specific entities (such as products or services) and their individual attributes (such as features or qualities) that users express opinions about. By distinguishing between these aspects ABSA enables a deeper understanding of customer feedback and provides more actionable insights. It has proven particularly valuable for businesses seeking to build fine-grained opinion summaries, offering a richer perspective on customer sentiment for applications in various sectors.

ABSA research focuses on four sentiment elements [2]: aspect term (*a*), aspect category (*c*), sentiment polarity (*p*), and opinion term (*o*). Given the review: “*The tea was very tasty*”, the corresponding sentiment elements are “*tea*”, “*drinks*”, “*positive*”, and “*tasty*”, respectively, as shown in Figure 1. Early ABSA research primarily focuses on *single* ABSA tasks identifying each sentiment element separately, such as aspect term extraction or aspect category detection. More recent work shifts towards *compound* ABSA tasks aiming at extracting two, three or four sentiment elements

^{*} Submitted version prior to peer review. Updated version accepted in *Information Fusion*: doi.org/10.1016/j.inffus.2025.103073.

^{*}Corresponding author

✉ jakusmid@kiv.zcu.cz (J. Šmíd); pkral@kiv.zcu.cz (P. Král)

ORCID(s): 0000-0002-4492-5481 (J. Šmíd); 0000-0002-3096-675X (P. Král)

together, making these tasks more complex and challenging. For example, end-to-end ABSA identifies the aspect terms and their corresponding sentiment polarities simultaneously.

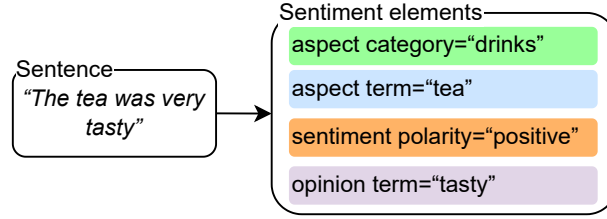


Figure 1: Example of a review with four ABSA sentiment elements.

Most existing research on ABSA has been conducted in a monolingual setting, with a strong focus on English. However, real-world users write reviews in many different languages [3]. A significant challenge is the limited availability of labelled datasets in languages other than English, and obtaining such annotated data is both expensive and time-consuming, especially for low-resource languages. This has led to the rise of cross-lingual ABSA as an important area of research. The goal of cross-lingual aspect-based sentiment analysis is to leverage labelled data from a *source language*, usually a resource-rich language like English, and transfer that knowledge to a *target language*, which typically lacks annotated data, as depicted in Figure 2. Due to the complexity of this task, cross-lingual ABSA often focuses on simpler tasks like aspect term extraction, aspect sentiment classification, and end-to-end ABSA, while monolingual studies have more recently tackled more intricate, compound tasks.

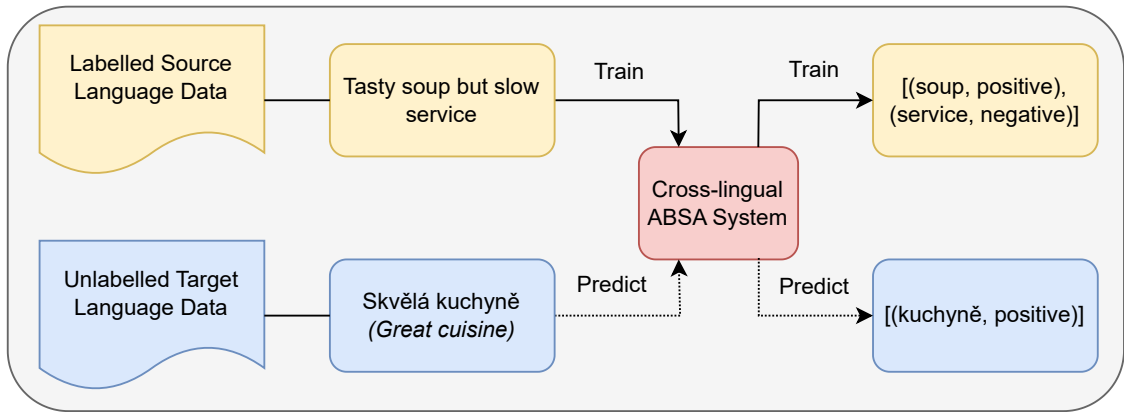


Figure 2: Visualization of cross-lingual ABSA system.

Several cross-lingual ABSA approaches have been explored, leveraging various forms of information fusion to address linguistic and cultural diversity across languages. Earlier methods combine machine translation with alignment algorithms to label translated data [4, 5] or employ cross-lingual word embeddings to enable models to switch languages by replacing the embedding layer [6, 7]. More recent work leverages multilingual pre-trained language models (mPLMs) such as multilingual BERT (mBERT) [8] and XLM-RoBERTa (XLM-R) [9], which have become standard tools for cross-lingual NLP tasks [10]. These models, trained on large multilingual corpora, fuse syntactic and semantic patterns across languages, enabling improved cross-lingual transfer. However, cross-lingual ABSA remains challenging, particularly in zero-shot settings where no labelled target language data is available, due to language-specific nuances such as slang, abbreviations, and unique aspect terms [11]. Recent advancements in this field often involve the fusion of mPLMs and machine translation with complementary techniques, such as alignment-free label projection [12], distillation on unlabelled target language data [12], contrastive learning [13], and dynamic weighted loss strategies [14], which collectively aim to bridge the linguistic gap and enhance model robustness in diverse languages. Recent NLP research has heavily focused on large language models (LLMs), but their application to cross-lingual ABSA remains limited to a few exploratory studies.

Existing ABSA surveys have a number of limitations. Earlier surveys [1, 15] take a broad approach to sentiment analysis without providing a detailed examination of ABSA. While later surveys focus on ABSA specifically [16, 17, 18, 2, 19], including modern neural network approaches, and some delve into specialized areas such as multimodal ABSA [20], they either completely overlook or provide limited coverage of cross-lingual approaches. This oversight leaves a gap in the systematic review of this increasingly important area. The primary aim of this paper is to bridge this gap by presenting a comprehensive and focused review of cross-lingual aspect-based sentiment analysis. We provide an in-depth discussion of ABSA tasks, methods for addressing them, available datasets, and general modelling paradigms. Furthermore, we review various techniques for cross-lingual transfer in ABSA, highlight existing challenges, and propose future research directions to advance the field.

The main contributions of this paper include:

- This paper is the first detailed survey focusing on cross-lingual aspect-based sentiment analysis, addressing a gap in the existing literature.
- This survey provides a comprehensive overview of the available datasets, ABSA tasks, and general modelling paradigms relevant to cross-lingual ABSA.
- This survey systematically reviews techniques for cross-lingual transfer in ABSA, highlighting advancements in multilingual pre-trained language models, machine translation, and other fusion-based approaches.
- This survey offers an in-depth analysis of solutions for specific cross-lingual ABSA tasks.
- This survey reviews related work in monolingual and multilingual ABSA, cross-lingual and multilingual sentiment analysis, as well as ABSA with LLMs, and examines their relevance to cross-lingual ABSA.
- This survey discusses the remaining challenges in cross-lingual ABSA and outlines potential future research directions to inspire advancements in this field.

The structure of the paper is as follows. Section 2 provides an overview of ABSA, covering its definitions, tasks, and available multilingual datasets. Section 3 discusses various modelling paradigms for ABSA. Section 4 focuses on techniques for cross-lingual transfer. Section 5 reviews existing solutions for different cross-lingual ABSA tasks in detail. Section 6 provides an overview of related research on innovative approaches in monolingual ABSA, multilingual ABSA, multilingual and cross-lingual sentiment analysis, and the application of large language models for ABSA, while also highlighting their connection to cross-lingual ABSA. Section 7 explores the limitations of current cross-lingual ABSA research and outlines potential directions for future studies. Finally, Section 8 summarizes the findings and contributions of this study to the field of cross-lingual ABSA.

2. ABSA Background

This section provides the background for aspect-based sentiment analysis, including definitions, tasks and datasets.

2.1. Definitions

Aspect-based sentiment analysis, also known as aspect-level sentiment analysis, performs fine-grained analysis, looking at opinions consisting of a sentiment and a target instead of looking at documents or sentences as a whole [21]. The target can be described by aspect term (a) or aspect category (c), while the sentiment can be described by sentiment polarity (p) or opinion term (o). These sentiment elements can be described as:

- **Aspect term** (a) is the specific target of the opinion directly mentioned in the text. For example, in the sentence “*The laptop’s battery lasts all day*”, the term “*battery*” is the aspect term. If the target is implied rather than explicitly stated (e.g. in “*It lasts forever!*”), the aspect term may be labelled as “*null*”.
- **Aspect category** (c) refers to a distinct attribute of an entity and is expected to belong to a predefined category set for a particular domain. For instance, in the electronics domain, categories like “*battery life*” or “*design*” could be used to evaluate different aspects of a smartphone.
- **Sentiment polarity** (p) refers to the emotional tone or attitude associated with an aspect category or aspect term, typically classified as positive, negative, or neutral. For example, the sentiment in “*The battery lasts all day*” is positive, while in “*The battery drains too fast*”, it is negative.

- **Opinion term** (*o*) represents the words or phrases the speaker uses to convey their sentiment about the target. In the sentence “*The laptop’s battery lasts all day*”, the term “*lasts all day*” is the opinion term describing the battery. Like aspect terms, opinion terms can be implicit, for example, in “*I am going back*”.

Given the sentiment elements definitions, we define ABSA as follows:

Aspect-based sentiment analysis involves identifying and extracting sentiment elements of interest from a given text, focusing on either individual sentiment elements or multiple elements that may exhibit dependency relationships. This process aims to uncover the sentiments associated with specific aspects of entities, enabling a comprehensive understanding of consumer opinions and facilitating informed decision-making in various applications.

We can also define cross-lingual ABSA as follows:

Cross-lingual aspect-based sentiment analysis aims to leverage labelled data from a resource-rich source language to enable aspect-based sentiment analysis in a low-resource target language. This approach facilitates the effective transfer of knowledge across languages, addressing the challenges posed by linguistic differences and enhancing the capacity for sentiment extraction regarding specific aspects in diverse languages.

In the literature, some terms are often used interchangeably but can have different meanings depending on the context. For instance, “*aspect*”, “*opinion target*”, “*entity*”, and “*target*” are commonly used to describe what the opinion is about. However, they can refer to either an aspect category or an aspect term, which can lead to confusion and incomplete literature reviews. In this survey, we use the most widely accepted terminology and clearly distinguish related concepts. As such, we differentiate between “*aspect term*” and “*aspect category*” while using “*target*” or “*aspect*” as general terms for an opinion target.

2.2. Tasks

Zhang et al. [2] classify ABSA tasks into two main categories, single and compound, depending on whether the output involves a single sentiment element or multiple linked elements. While single tasks are generally simpler, they fail to capture more comprehensive opinions at the aspect level. Consequently, recent research has shifted towards tackling compound tasks, which provide a more complete understanding of opinions. This shift is significant as it allows for a more nuanced analysis of opinions, leading to more accurate sentiment analysis results.

Figure 3 illustrates the taxonomy of ABSA tasks, highlighting representative works in cross-lingual settings. Additionally, the figure identifies tasks for which no references are available, indicating gaps in cross-lingual ABSA research and highlighting opportunities for future exploration. Table 1 presents the input and output examples for each task. Below is a brief overview of the key ABSA tasks:

- **Aspect term extraction** (ATE) focuses on identifying specific aspect terms that are the target of opinions expressed in the text.
- **Aspect category detection** (ACD) aims to determine the broad categories under which the aspects discussed in the text fall.
- **Opinion term extraction** (OTE) identifies opinion expressions towards specific aspects. It includes two sub-tasks. Aspect opinion co-extraction (AOCE) extracts aspect and opinion terms without linking them. However, it does not link the aspect terms and opinions together, meaning it is still considered a single ABSA task. The second task is target-oriented opinion word extraction (TOWE), which aims to extract the corresponding opinion terms given a specific aspect term.
- **Aspect sentiment classification** (ASC) predicts the sentiment polarity for a specific aspect, which may be an aspect term, category, or a combination of both.
- **Aspect-opinion pair extraction** (AOPE) focuses on extracting both aspect terms and their corresponding opinion expressions from the text.
- **End-to-end ABSA** (E2E-ABSA) extracts both aspect terms and their associated sentiment polarities.

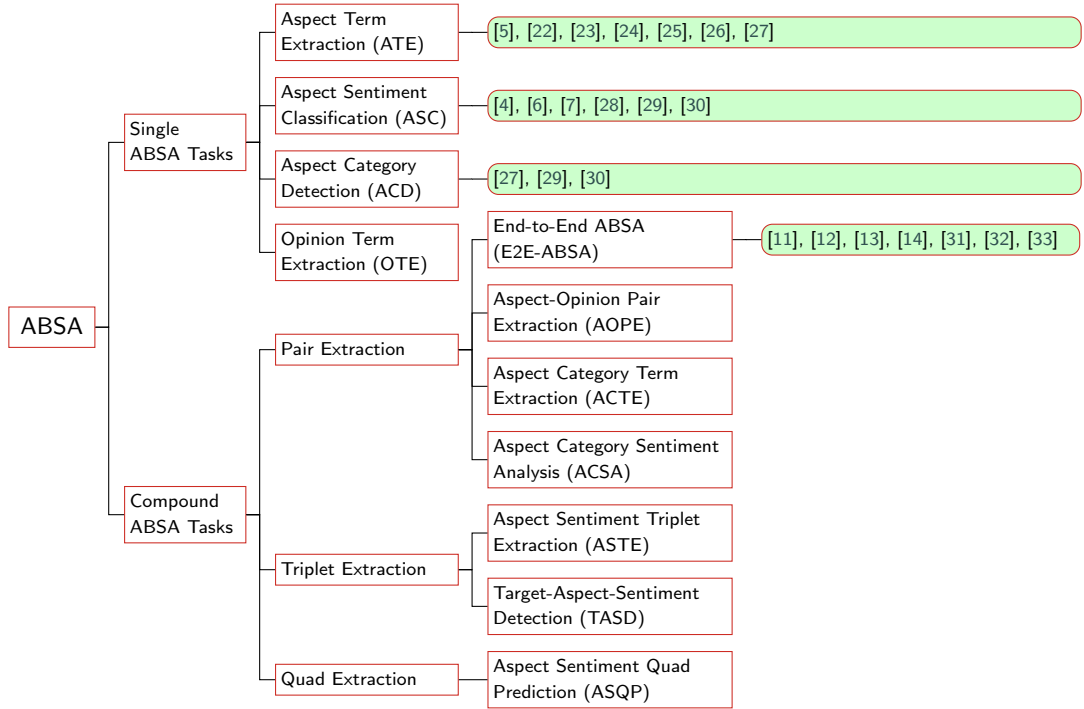


Figure 3: Taxonomy of ABSA tasks, highlighting representative papers addressing each task in cross-lingual settings. Tasks without references indicate gaps in cross-lingual research.

- **Aspect category term extraction (ACTE)** aims to extract aspect terms alongside their respective aspect categories.
- **Aspect category sentiment analysis (ACSA)** involves detecting aspect categories discussed in the text and predicting their corresponding sentiment polarities.
- **Aspect sentiment triplet extraction (ASTE)** extracts (aspect term, opinion term, sentiment polarity) triplets.
- **Target-aspect-sentiment detection (TASD)** extracts (aspect term, aspect category, sentiment polarity) triplets for a given sentence. It is also referred to as aspect-category-sentiment detection (ACSD).
- **Aspect sentiment quad prediction (ASQP)** focuses on extracting all (aspect term, opinion term, aspect category, sentiment polarity) quadruplets from a sentence.

Given the difficulty of cross-lingual ABSA compared to monolingual ABSA and also much less cross-lingual aspect-based sentiment analysis research, most of the cross-lingual ABSA work focuses on simple tasks such as aspect term extraction and aspect sentiment classification, and from compound tasks, only on end-to-end ABSA. This lack of research on compound ABSA tasks in cross-lingual settings presents one of the key challenges and limitations of current ABSA research.

2.3. Datasets

Annotated datasets are vital for the development of ABSA methods. While most datasets are in English, this section focuses on commonly used multilingual datasets for cross-lingual ABSA and briefly mentions widely used monolingual datasets. Table 2 summarizes the key multilingual datasets, including languages, domains, and annotated sentiment elements.

The SemEval-2016 dataset [3] contains real-world restaurant reviews and is one of the most widely used in cross-lingual ABSA. It builds on earlier shared task datasets from SemEval-2014 [38] and SemEval-2015 [39], expanding from English to include Spanish, French, Dutch, Russian, and Turkish. This dataset features aspect terms, categories,

Table 1

An overview of the input and output examples for each ABSA task considering an input sentence s : “The tea was tasty, but the waiter was rude.”.

Task	Example input	Output	Example output
<i>Single tasks</i>			
Aspect term extraction (ATE)	s	$\{a\}$	$\{\text{"tea"}, \text{"service"}\}$
Aspect category detection (ACD)	s	$\{c\}$	$\{\text{"food"}, \text{"service"}\}$
Aspect opinion co-extraction (AOCE)	s	$\{a\}, \{o\}$	$\{\text{"tea"}, \text{"waiter"}\}, \{\text{"tasty"}, \text{"rude"}\}$
Target-oriented opinion word extraction (TOWE)	$s, \text{"tea"}$	o_1	"tasty"
	$s, \text{"waiter"}$	o_2	"rude"
Aspect sentiment classification (ASC)	$s, \text{"tea"}$	p_1	"positive"
	$s, \text{"waiter"}$	p_2	"negative"
<i>Compound tasks</i>			
Aspect-opinion pair extraction (AOPE)	s	$\{(a, o)\}$	$\{(\text{"tea"}, \text{"tasty"}), (\text{"waiter"}, \text{"rude"})\}$
End-to-end ABSA (E2E-ABSA)	s	$\{(a, p)\}$	$\{(\text{"tea"}, \text{"positive"}), (\text{"waiter"}, \text{"negative"})\}$
Aspect category term extraction (ACTE)	s	$\{(a, c)\}$	$\{(\text{"tea"}, \text{"food"}), (\text{"waiter"}, \text{"service"})\}$
Aspect category sentiment analysis (ACSA)	s	$\{(c, p)\}$	$\{(\text{"food"}, \text{"positive"}), (\text{"service"}, \text{"negative"})\}$
Aspect sentiment triplet extraction (ASTE)	s	$\{(a, o, p)\}$	$\{(\text{"tea"}, \text{"tasty"}, \text{"positive"}), (\text{"waiter"}, \text{"rude"}, \text{"negative"})\}$
Target-aspect-sentiment detection (TASD)	s	$\{(a, c, p)\}$	$\{(\text{"tea"}, \text{"food"}, \text{"positive"}), (\text{"waiter"}, \text{"service"}, \text{"negative"})\}$
Aspect sentiment quad prediction (ASQP)	s	$\{(a, c, o, p)\}$	$\{(\text{"tea"}, \text{"food"}, \text{"tasty"}, \text{"positive"}), (\text{"waiter"}, \text{"service"}, \text{"rude"}, \text{"negative"})\}$

Table 2

An overview of common multilingual ABSA benchmark datasets. Languages are denoted using ISO 639-1 codes. *The languages in the MultiAspectEmo dataset, except for Polish, are translations of the Polish reviews.

Dataset	Languages	Major Domains	Annotations
SemEval-2016 [3]	ar, en, es, fr, nl, ru, tr, zh	restaurants, hotels, electronics	a, c, p
USAGE [34]	en, de	products	a, p
OpeNER [35]	en, es	hotels	a, p
[36]	en, kr	automotive	a, p
MultiAspectEmo* [32]	cs, en, es, fr, nl, pl	school, medicine, hotels, products	a, p
PanoSent [37]	en, es, zh	more than 100 domains	<i>multimodal annotations</i>

and sentiment polarities, enabling various compound ABSA tasks. However, it lacks opinion term annotations, which were later introduced in other works [40, 41, 42, 43, 44], though only for English. In addition to the restaurant domain, the SemEval-2016 dataset includes hotel reviews in Arabic and electronics reviews in Chinese. Introduction of a new Czech dataset in the SemEval-2016 format, expanded and reannotated based on prior datasets [45, 46], enables future cross-lingual comparisons [47].

Other notable multilingual datasets include the USAGE corpus [34], featuring English and German product reviews with annotated aspect terms and sentiment polarities, and the OpeNER dataset [35], which contains hotel reviews in English and Spanish. A dataset in the automotive domain for English and Korean contains annotations of aspect terms and sentiment polarities [36]. The MultiAspectEmo dataset [32] contains original reviews in Polish in four domains, along with translations into English, Czech, Spanish, French, and Dutch. The PanoSent dataset [37] covers more than 100 domains for multimodality in English, Chinese, and Spanish.

The Korean dataset [30], focused on the restaurant domain, follows a format similar to the SemEval-2014 dataset [38]. Notably, a portion of this dataset is a direct translation of the SemEval-2014 dataset into Korean.

Table 3

Overview of various modelling paradigms used in cross-lingual ABSA, the tasks they address, and representative papers.

Modelling Paradigm	Task	Sources
Sequence-Level Classification	Aspect Sentiment Classification	[4, 6, 7, 28, 29, 30]
	Aspect Category Detection	[27, 29, 30]
Token-Level Classification	Aspect Term Extraction	[5, 22, 23, 24, 25, 26, 27]
	End-to-End ABSA	[11, 12, 13, 14, 31, 32]
Sequence-to-Sequence Modelling	End-to-End ABSA	[33]

For monolingual datasets, a Hindi dataset with electronics reviews is available [48], as well as a Japanese dataset for hotel reviews annotated with aspect categories and sentiment polarities [49]. For Chinese, there is a dataset for ASC in QA-style reviews [50], and ASAP [51], a large-scale dataset in the restaurant domain. Other English datasets include SentiHood [52], MAMS [53], and ARTS [54].

One key limitation of multilingual ABSA datasets is their limited language coverage and the absence of opinion term annotations. This gap has prevented the exploration of ABSA tasks involving opinion terms across multiple languages.

2.4. Evaluation Metrics

For evaluation, an exact match is the dominant method across ABSA datasets, where a prediction is correct only if all elements match annotations. Standard classification metrics like accuracy, precision, recall, and F1 score are then used to assess model performance.

3. Modelling Paradigms

Several paradigms are commonly used to solve ABSA tasks [2], including sequence-level classification, token-level classification, and sequence-to-sequence modelling. Each paradigm represents a general computational framework designed to handle a specific input-output format, allowing them to be applied across various tasks. Beyond these unified paradigms, which address tasks in an end-to-end manner, more complex ABSA tasks may require the pipeline paradigm, where multiple models are connected sequentially to produce the final prediction. However, most cross-lingual ABSA research primarily relies on sequence-level and token-level classification paradigms; therefore, this section will discuss them in greater detail compared to other paradigms. Table 3 presents representative cross-lingual ABSA works for different tasks and modelling paradigms.

In this section, we denote the dataset for a given ABSA task as $\mathcal{D} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{|\mathcal{D}|}$, where $\mathbf{x}_i = \{x_1, x_2, \dots, x_n\}$ represents the i -th input sentence consisting of n tokens, and \mathbf{y}_i is the corresponding label.

3.1. Sequence-Level Classification

In sequence-level classification, a model typically consists of an encoder $\mathbf{Enc}(\cdot)$ that processes input text \mathbf{x} to extract features specific to the task. The extracted features are then passed through a classifier $\mathbf{CLS}(\cdot)$ to produce the label \mathbf{y} as

$$\mathbf{y} = \mathbf{CLS}(\mathbf{Enc}(\mathbf{x})). \quad (1)$$

With advancements in deep learning, the encoder architecture can range from convolutional networks (CNNs) [55], recurrent neural networks (RNNs) like LSTMs [56], and Transformers [57], enabling the capture of rich contextual information. For tasks like aspect sentiment classification, where inputs include both a sentence and an associated aspect term or category, the encoder must model not only the individual inputs but also their interactions. The classifier component is commonly designed as a multi-layer perceptron, often incorporating a pooling mechanism to derive the final prediction.

Aspect category detection is often treated as a multi-label classification problem in this paradigm, where each aspect category is treated as a separate label. Sequence-level classification is frequently used for aspect sentiment

classification, and this paradigm is commonly explored in cross-lingual settings. Figure 4 illustrates the sequence-level classification paradigm for aspect sentiment classification.

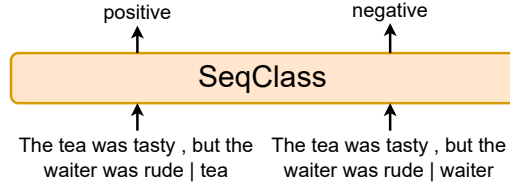


Figure 4: Sequence-level classification paradigm for the aspect sentiment classification task.

3.2. Token-Level Classification

In contrast to sequence-level classification, which assigns a single label to the entire input, token-level classification (also known as sequence labelling or sequence tagging) assigns a label to each individual token in the input sequence. Similar to sequence-level classification, token-level classification begins with encoding the input text into contextualized features using an encoder $\mathbf{Enc}(\cdot)$. A decoder $\mathbf{Dec}(\cdot)$ then predicts labels y_1, y_2, \dots, y_n for each token x_1, x_2, \dots, x_n in the input \mathbf{x} as

$$y_1, y_2, \dots, y_n = \mathbf{Dec}(\mathbf{Enc}(x_1, x_2, \dots, x_n)). \quad (2)$$

The decoder is often implemented as a multi-layer perceptron with a softmax layer or with conditional random fields (CRF) [58]. Various tagging schemes, such as the BIOES scheme (Beginning, Inside, Outside, Eend, Singleton) [59] or the BIO scheme, can also be employed.

Aspect term extraction problem is commonly framed as a token-level classification task, as aspect terms are typically words or phrases within a sentence. This paradigm is also used for tasks like aspect opinion co-extraction and end-to-end ABSA, often with tagging schemes like BIOES. Since aspect term extraction and end-to-end ABSA are commonly examined in cross-lingual research, token-level classification is a popular approach in cross-lingual ABSA studies. Figure 5 shows an example of a token-level classification paradigm for aspect term extraction.

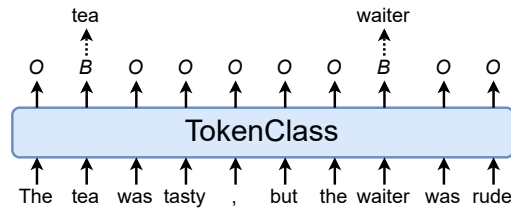


Figure 5: Token-level classification paradigm for the aspect term extraction task.

3.3. Sequence-to-Sequence Modelling

The sequence-to-sequence (Seq2Seq) paradigm processes an input sequence $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ and generates an output sequence $\mathbf{y} = \{y_1, y_2, \dots, y_m\}$. Although commonly used in machine translation tasks, it is also applied in ABSA tasks, where it can generate label sequences or extract sentiment-related elements from input text. For instance, in the aspect term extraction task, an input \mathbf{x} like “*The steak with potatoes was delicious*” could yield an output \mathbf{y} like “*steak with potatoes*”.

The sequence-to-sequence paradigm typically employs an encoder-decoder model, such as the Transformer, where the encoder $\mathbf{Enc}(\cdot)$ extracts contextualized features from the input, and the decoder $\mathbf{Dec}(\cdot)$ generates tokens based on the encoded input and previously generated tokens as

$$y_1, y_2, \dots, y_m = \mathbf{Dec}(\mathbf{Enc}(x_1, x_2, \dots, x_n)). \quad (3)$$

With the advent of large language models, which are typically decoder-only models, this paradigm has gained prominence. In this case, the decoder generates token y_i based only on previously generated tokens as

$$y_i = \text{Dec}(y_1, y_2, \dots, y_{i-1}). \quad (4)$$

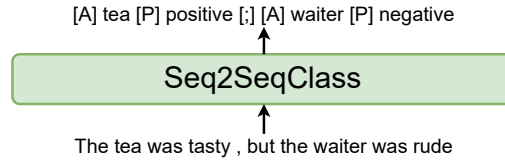


Figure 6: Sequence-to-sequence modelling paradigm for end-to-end ABSA.

Although Seq2Seq modelling can be used to solve individual ABSA tasks (e.g. aspect term extraction), it is mainly employed for compound ABSA tasks. Figure 6 illustrates the use of sequence-to-sequence modelling for end-to-end ABSA. This paradigm has gained significant traction in recent years within monolingual ABSA research [44, 60, 61, 62, 63, 64], including studies with LLMs [62, 65, 66]. However, its potential in cross-lingual ABSA remains largely unexplored, with only one preliminary findings reported in a single study [33] using fine-tuned LLMs.

3.4. Pipeline Method

The pipeline approach is frequently used for complex, compound ABSA tasks. It involves connecting multiple models sequentially, where the output of one model becomes the input for the next, as shown in 7. For instance, in the AOPE problem, an aspect term extraction model identifies aspect terms, which are then passed to a subsequent model to extract corresponding opinion terms. However, this method is prone to error propagation, where mistakes made by earlier models can negatively impact the overall performance. Despite its utility in handling complex tasks, the pipeline method has not been widely adopted in cross-lingual ABSA research.

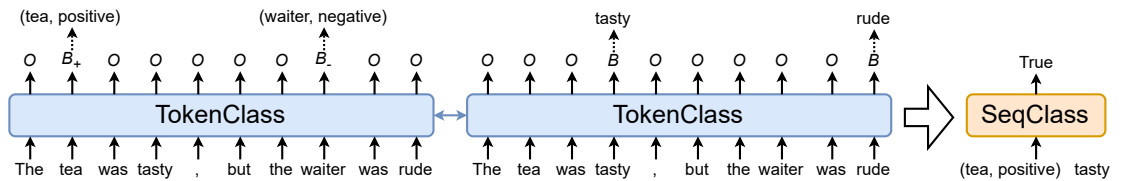


Figure 7: Pipeline method for aspect sentiment triplet extraction.

4. Cross-lingual Transfer

In cross-lingual ABSA, cross-lingual transfer methods facilitate the fusion of linguistic knowledge between source and target languages, reducing the gap and enabling effective sentiment analysis across diverse linguistic contexts. The three primary approaches to achieving this transfer are the fusion of cross-lingual word embeddings, the integration of machine translation techniques, and the use of multilingual pre-trained language models.

4.1. Cross-lingual Word Embeddings

Pre-trained word embeddings, such as Word2Vec [67], FastText [68], and GloVe [69], have been widely used in NLP to capture semantic relationships between words. These embeddings represent words as dense, continuous vectors in a high-dimensional space, capturing semantic relationships and syntactic patterns. Unlike sparse one-hot encodings, word embeddings leverage context from large corpora to create vectorized representations, where semantically similar words are close to each other in the embedding space.

Word2Vec, one of the pioneering models, learns embeddings using *skip-gram* or *continuous bag-of-words* (CBOW) techniques, predicting the context of a word or vice versa. FastText extends Word2Vec by incorporating subword

information, which improves its ability to handle rare and morphologically complex words. GloVe, on the other hand, focuses on word co-occurrence statistics to capture the global context. These pre-trained embeddings have been widely adopted in various NLP tasks, including text classification, sentiment analysis, and machine translation, due to their ability to improve performance with minimal training data.

These embeddings are widely used in tasks like text classification, sentiment analysis, and machine translation. For classifiers like support vector machines (SVMs), word embeddings are typically averaged across all words in a sentence, as shown in Figure 8. While effective, this method may lose some of the embeddings' rich information. Neural network models, such as CNNs and LSTMs, can process sequences of word vectors more effectively, as illustrated in Figure 9, which shows a simple CNN architecture.

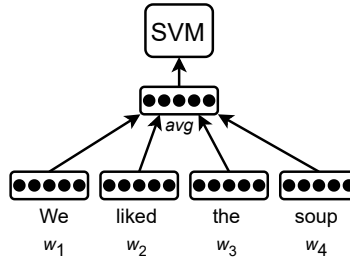


Figure 8: Example of averaging word embeddings for each word w_n to obtain a single vector representing a whole sentence, which is then passed to a classifier such as SVM.

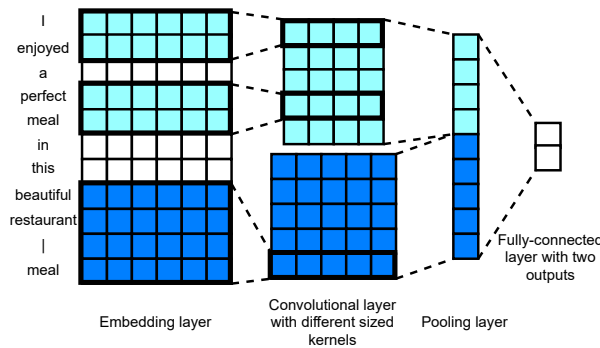


Figure 9: Example of a convolutional neural network with input embeddings.

4.1.1. Bilingual Word Embeddings

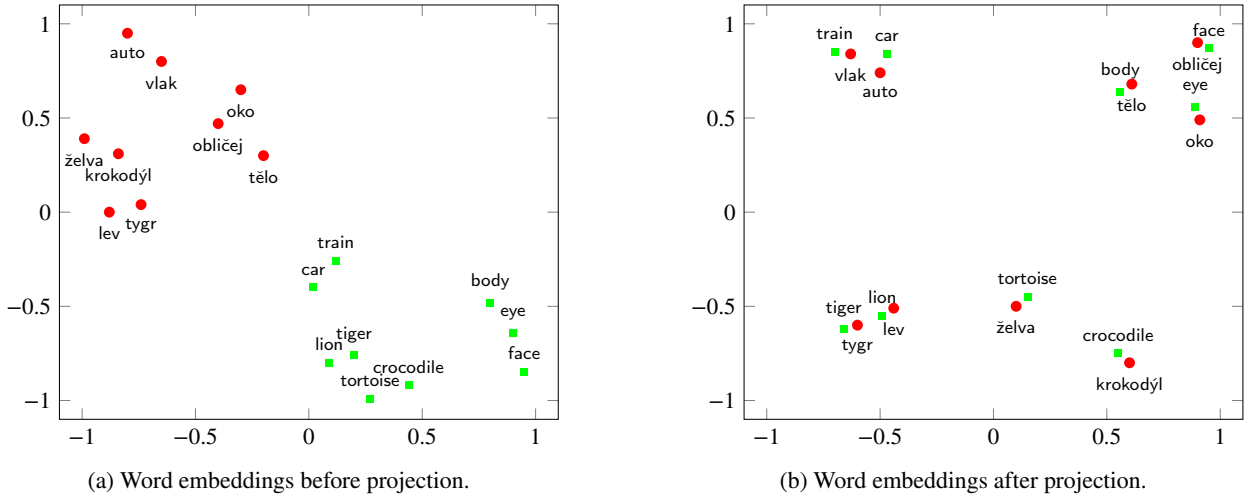
In cross-lingual tasks, bilingual word embeddings (BWEs) extend monolingual embeddings to multilingual scenarios. For instance, the *bilingual skip-gram algorithm* [70] modifies the traditional skip-gram model by incorporating bilingual objectives, ensuring that words in parallel sentences have similar embeddings across languages.

Barnes et al. [6] train bilingual embeddings on the English-Spanish Europarl corpus [71]. Similarly, Akhtar et al. [7] train bilingual word embeddings on a parallel English-Hindi and English-French corpora. They propose an approach to deal with out-of-vocabulary (OOV) words, i.e. words whose representations are missing in one of the bilingual embeddings vocabularies. Instead of generating random vectors or omitting these OOV words, which are common approaches to deal with them, they instead translate missing words into the other language, e.g. English, and use the word embedding corresponding to the translated word. Finally, they feed the word embeddings to the LSTM classifier.

4.1.2. Word Embeddings Alignment

Another common approach involves aligning monolingual embeddings from different languages into a shared space, enabling cross-lingual transfer. The alignment can be achieved by learning a transformation matrix \mathbf{W} that maps embeddings from one language into the space of another, effectively bridging the gap between languages. For example,

given pre-trained word embeddings \mathbf{X}^s for the source language s and \mathbf{X}^t for the target language t , a transformation matrix $\mathbf{W}^{s \rightarrow t}$ is learned to align these embeddings, such that the embeddings in language s are mapped to the same space as those in language t , ideally preserving the semantic relationships between words across languages, as shown in Figure 10. This can be formulated as minimizing the distance between corresponding word pairs in a bilingual dictionary, such as using regression, orthogonal or canonical methods [72].



To solve the alignment problems during translation, Lambert [4] translates entire sentences while preserving opinionated segments through reordering constraints, ensuring that text between segment boundaries remains intact and unaltered by external text. They employ the Moses SMT toolkit [74] for this translation, utilizing zone and wall tags to enforce these constraints and enabling the passage of markup to track segment boundaries and their corresponding identifiers and polarity labels, as shown in Figure 11.

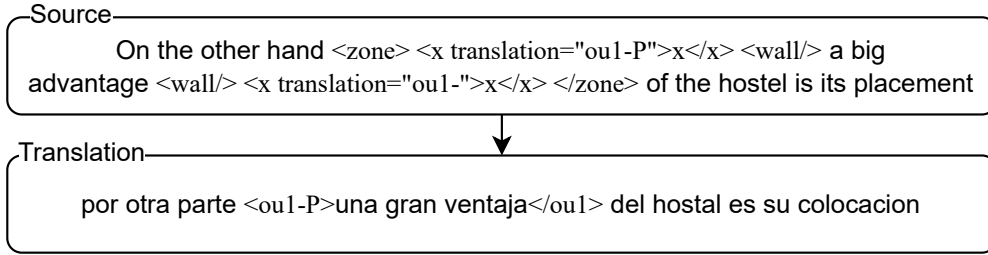


Figure 11: Source text that includes reordering constraint markup along with the code to transfer tags and their corresponding translation, as proposed by Lambert [4].

Traditionally, the source and translated target sentences are processed to generate word alignment links with an alignment tool indicating which words in the target language correspond to words in the source sentence. Although aligning word-level labels offers a straightforward method for token-level label projection, it is vulnerable to issues such as word order changes and alignment gaps. To enhance alignment quality, Li et al. [11] propose a span-to-span mapping strategy, depicted in Figure 12, that transforms word-based alignments into span-based alignments, allowing for more robust label propagation across aligned spans while mitigating potential problems with word order changes.

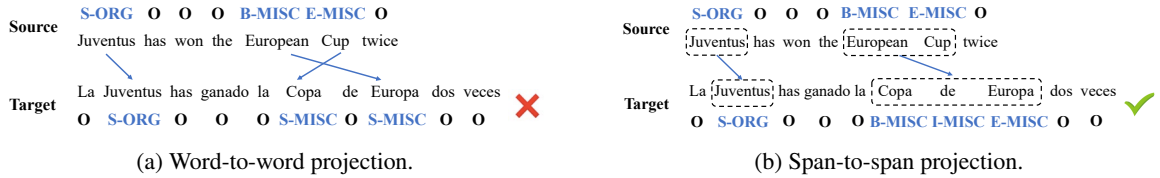


Figure 12: Example sentence with different label projection strategies with BIOES as the tagging scheme. The span-to-span strategy proposed by Li et al. [11].

Zhang et al. [12] introduce an alignment-free label projection method to generate pseudo-labelled data in the target language. Initially, they mark each aspect term in the source sentence with unique symbols before translating it. Once the translation is completed, they extract the spans marked with these symbols and align them with the corresponding aspect terms from the source to recover the aspect boundaries and project the sentiment labels. This approach ensures that even if the order of aspect terms changes during translation, they can still be matched to the appropriate labels; however, they also account for cases where symbols may be lost during translation by filtering out sentences missing these markers. Additionally, the authors propose constructing bilingual sentences by switching aspect terms between the source sentence and its translation. This results in two datasets: one with aspect terms in the target language and the other with aspect terms in the source language. The corresponding label sequences are adjusted accordingly, creating code-switched datasets for further analysis. Figure 13 depicts the proposed method.

Nam [30] utilizes Google Translate to create a machine-translated version of the English SemEval-2014 dataset in Korean. Since the tasks addressed are ASC and ACD, no alignment of aspect terms and opinions is required. In contrast, other work [32] leverages the DeepL API to construct a multilingual dataset by translating Polish data into five different languages, employing XLM markups to align aspect terms effectively.

4.3. Multilingual Pre-trained Language Models

Multilingual pre-trained language models, such as mBERT [8], XLM-R [9] and mT5 [75], have become popular for cross-lingual NLP tasks [10]. These models are pre-trained on large multilingual corpora, enabling them to capture cross-lingual syntactic and semantic patterns.

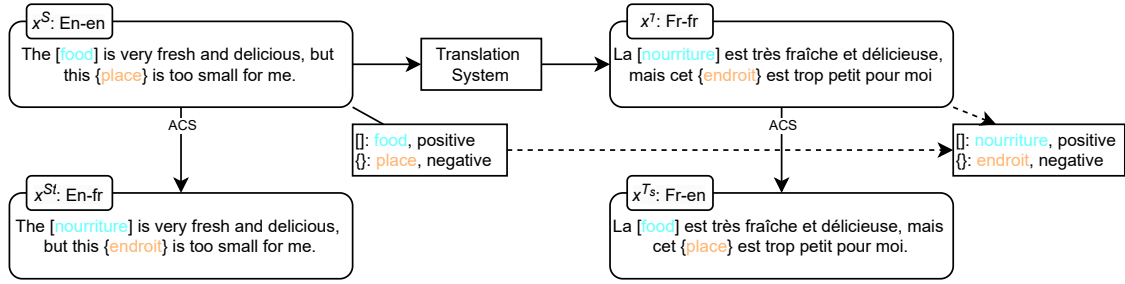


Figure 13: Example of the aspect code-switching technique (lower part) and alignment-free label projection method (upper part) proposed by Zhang et al. [12].

Fine-tuning these models on labelled data in the source language allows *zero-shot* transfer, meaning the model can perform well in the target language without any explicit training data in that language. This approach has been highly effective in many cross-lingual tasks [76, 77, 78].

Most mPLMs are based on the Transformer architecture [57], which utilizes self-attention mechanisms to capture relationships between words, regardless of their position in the sequence. Figure 14 illustrates the architecture, with some models (like mBERT and XLM-R) using only the encoder for classification tasks, while others (like mT5) use both encoder and decoder components for tasks involving sequence generation.

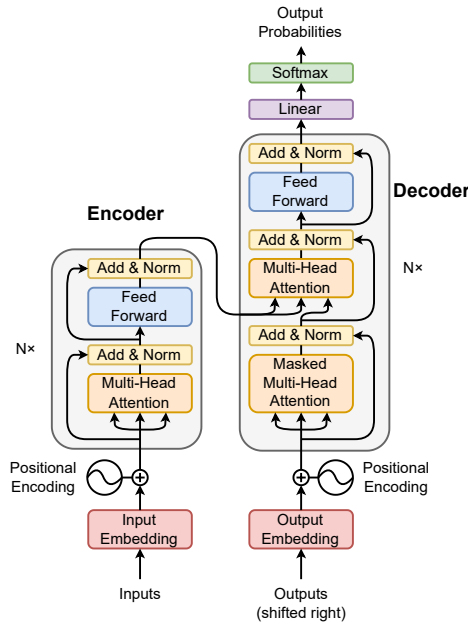


Figure 14: Transformer architecture [57].

In most cases, models employ subword tokenizers to handle out-of-vocabulary words. Rather than processing entire words, subword tokenizers divide words into smaller units, allowing models to effectively handle rare or unknown words. Another key advantage of mPLMs is that they produce contextualized word embeddings, which capture the meaning of words based on their context. This capability is a significant improvement over static embeddings from models like Word2Vec, FastText, or GloVe, which return the same vector for a word regardless of context – leading to identical vectors for polysemous words like “*crane*” (whether referring to a bird or a machine) – mPLMs generate different vectors based on the word’s context.

Phan et al. [27] explore zero-shot capabilities with mBERT and XLM-R across various source and target language combinations for ACD and ATE tasks, showing that XLM-R outperforms mBERT in most scenarios. Similarly, Dang

Table 4

Overview of key cross-lingual ABSA works, organized by year. The **Approach** column categorizes methods as classical machine learning (ML), neural networks (NN), or Transformers (T). The **Transfer Method** column specifies the approach for cross-lingual transfer: machine translation (MT), cross-lingual/bilingual word embeddings (WE), or multilingual pre-trained language models (mPLMs). The **Languages** column lists the languages using ISO 639-1 codes. *This study primarily utilizes English-centric LLMs; however, these models also incorporate data from languages other than English during their initial pre-training phase.

Source	Year	Approach	Method (Model)	Transfer Method	Data Domain	ABSA Tasks	Languages
[22]	2014	ML	LDA-based	MT	Hotels, products	ATE	de, en, es, fr, it, nl, zh
[23]	2015	ML	CRF	MT	Products	ATE	de, en
[4]	2015	ML	SVM	MT	Hotels	ASC	en, es
[5]	2015	ML	CRF	MT	Electronics, cars	ATE	en, zh
[6]	2016	ML, NN	SVM, LSTM	MT, WE	Hotels	ASC	en, es
[24]	2018	NN	LSTM	WE	Restaurants, laptops	ATE	en, hi
[7]	2018	NN	LSTM	WE	Restaurants, laptops	ASC	en, fr, hi
[25]	2018	NN	RNN	WE	Restaurants	ATE	en, es, fr
[26]	2019	NN	CNN	WE	Restaurants	ATE	en, es, nl, ru, tr
[11]	2020	T	mBERT, XLM-R	MT, mPLMs	Restaurants	E2E-ABSA	en, es, fr, nl, ru, tr
[27]	2021	T	mBERT, XLM-R	mPLMs	Restaurants	ACD, ATE	en, es, fr, nl, ru, tr
[31]	2021	NN, T	CNN, mBERT	mPLMs	Restaurants	E2E-ABSA	en, es, fr, nl
[28]	2021	T	BERT	MT	Restaurants, laptops	ASC	en, es, ru, zh
[12]	2021	T	mBERT, XLM-R	MT, mPLMs	Restaurants	E2E-ABSA	en, es, fr, nl, ru
[32]	2022	T	XLM-R, LaBSE	mPLMs	Restaurants, school, medicine, hotel, products	E2E-ABSA	cs, en, es, fr, nl, pl
[29]	2023	T	mBERT, XLM-R	mPLMs	Restaurants	ACD, ASC	en, es, fr, nl, ru
[13]	2023	T	mBERT	MT, mPLMs	Restaurants	E2E-ABSA	en, es, fr, nl, ru
[14]	2024	T	mBERT, XLM-R	MT, mPLMs	Restaurants	E2E-ABSA	en, es, fr, nl, ru
[30]	2024	T	mBERT, XLM-R, LaBSE	mPLMs	Restaurants	ACS, ASC	en, kr
[33]	2024	T	LLMs	mPLMs*	Restaurants	E2E-ABSA	en, es, fr, nl, ru

Van Thin and Nguyen [29] demonstrate that XLM-R performs better in ACD and ASC tasks while also highlighting the benefits of fine-tuning the models on multiple source languages rather than just one.

Nam [30] leverages mBERT and XLM-R to generate pseudo-labelled data for Korean in ACD and ASC tasks, incorporating LaBSE [79] to filter out low-quality labels. Similarly, different study [32] employs XLM-R and LaBSE for the E2E-ABSA task across various source and target language combinations.

Despite the promise of zero-shot approaches for cross-lingual ABSA, several challenges remain. Language-specific knowledge is vital, as user-generated content often includes abbreviations, slang, or informal language. Aspect and opinion terms can differ significantly across languages. However, zero-shot methods rely heavily on pre-trained models, where low-resource languages are often underrepresented [9, 80]. This underrepresentation limits the model's ability to capture and transfer language-specific nuances effectively. To mitigate these limitations, some studies combine machine translation with mPLMs for E2E-ABSA, incorporating additional techniques such as parameter warm up [11], distillation on target language data [12], contrastive learning [13], and handling imbalanced classes [14].

Other research bypasses fine-tuning mPLMs directly and instead leverages their ability to generate contextualized word embeddings. For instance, Sattar et al. [31] use these context-aware embeddings from mBERT as input to a CNN, facilitating a more nuanced understanding of word meanings.

5. Cross-lingual ABSA Tasks

This section discusses ABSA tasks explored in cross-lingual settings in more detail. For the cross-lingual E2E-ABSA task, we include tables with results as consistent datasets, models, and languages used in most of the works allow for a direct comparison of results. In contrast, methods are evaluated across diverse datasets and languages for other tasks, making it impractical to present results in a unified table. Therefore, we present an overview of the cross-lingual ABSA work alongside the solved tasks, cross-lingual transfer used, languages, dataset domains, and models in Table 4.

5.1. Aspect Term Extraction

Aspect term extraction is a core task in ABSA, focusing on identifying specific aspect expressions within a given text that users express opinions about. For instance, in the sentence “*The tea is great, but the ambience is disappointing*”, the aspect terms “*tea*” and “*ambience*” would be extracted as they represent the subjects of the user’s opinions. Aspect term extraction is often framed as a token-level classification problem, where each token in the sentence is classified, as the aspect terms are usually individual words or short phrases.

Zhou et al. [5] employ machine translation and co-training of two labelling models based on CRF to transfer knowledge between English and Chinese. In this approach, each model annotates unlabelled datasets, selects the most confident examples, and adds them to the labelled dataset to continue training. Similarly, Klinger and Cimiano [23] use CRF combined with machine translation to address ATE between English and German. Akhtar et al. [24] employ LSTM with bilingual embeddings to tackle ATE in English and Hindi. Lin et al. [22] propose an unsupervised LDA-based model to align aspects bilingually by assuming that reviews in different languages share the same topic distribution, using a bilingual dictionary to translate words and select the most suitable translation, enabling cross-lingual topic alignment without the need for parallel corpora.

Wang and Pan [25] examine English, French, and Spanish and introduce a transition-based mechanism that processes one word at a time, forming a series of configurations that represent the status of the entire sentence. Each configuration is represented as a continuous feature vector, which is aligned from different languages into a shared space through an adversarial network. The transition-based adversarial network consists of a generator that produces language-invariant configuration features and a discriminator that distinguishes configurations between source and target languages. Bilingual word embeddings are the input to the network.

Jebbara and Cimiano [26] investigate multilingual word embeddings for English, Spanish, Dutch, Russian, and Turkish using CNNs. They experiment with various combinations of source and target languages and find that cross-lingual learning from multiple source languages enhances performance. Similarly, Phan et al. [27] explore mPLMs, specifically mBERT and XLM-R, in zero-shot settings for English, Spanish, French, Dutch, Russian, and Turkish. Their findings indicate that XLM-R typically outperforms mBERT across different source and target language combinations.

5.2. Aspect Sentiment Classification

Aspect sentiment classification, often referred to as aspect-based or aspect-level sentiment classification, focuses on determining the sentiment polarity associated with a specific aspect within a sentence. The aspect can be either an aspect term or an aspect category, leading to two distinct ASC tasks: aspect term-based sentiment classification and aspect category-based sentiment classification. While there are minor distinctions between the two (such as utilizing the position information of a given aspect term from the sentence), the primary research question remains consistent across both: how to effectively leverage the relationship between the aspect (whether a term or category) and the surrounding sentence context to classify sentiment accurately. It is possible to address both subtasks simultaneously, employing the same model to handle them seamlessly.

Lambert [4] uses constraint SMT (see Section 4.2 for more details) alongside SVM to perform aspect term-based sentiment classification in Spanish and English, adding the tokens (unigrams) of aspect terms as additional features to the classifier.

Barnes et al. [6] compare several methods for aspect term-based sentiment classification in Spanish and English with the SVM classifier. In zero-shot settings, they train a translation matrix to map two monolingual vector spaces into one. Additionally, they train bilingual word embeddings and employ stacked denoising autoencoders to encode the parallel sentences into a common latent space. Finally, they experiment with machine translation. They find that the machine translation approach performs the best, likely due to its use of high-quality, in-domain data. Additionally, they evaluate LSTM but demonstrate that SVM is more efficient in their setup.

Akhtar et al. [7] utilize bilingual word embeddings and various hand-crafted features as the input to the bidirectional LSTM classifier for aspect term-based sentiment classification in English, Hindi, and French. They employ concatenation to fuse the aspect information with the sentence context. Wu et al. [28] use machine translation to build a bilingual lexicon to explore aspect category-based sentiment classification in Chinese, Russian, English, and Spanish. They introduce a reinforced transformer based on pre-trained English BERT [8] with cross-lingual distillation to model aspect representations and sentence representations.

Dang Van Thin and Nguyen [29] explore mPLMs, specifically mBERT and XLM-R, to tackle aspect category-based sentiment classification across English, Spanish, French, Dutch, and Russian with joint training on data from

Table 5

Example of the BIOES and BIO tagging scheme for the end-to-end ABSA task.

	The	steak	with	potatoes	was	delicious	,	but	the	service	was	horrible	.
BIOES	O	B-POS	I-POS	E-POS	O	O	O	O	O	S-NEG	O	O	O
BIO	O	B-POS	I-POS	I-POS	O	O	O	O	O	B-NEG	O	O	O

multiple languages. Their models output a one-hot vector with a size equal to the number of categories, where each value represents the sentiment polarity of a category corresponding to its position in the vector.

Nam [30] introduces a framework that fine-tunes multilingual BERT-base models on the Korean translation of the English SemEval-2014 dataset for the aspect category-based sentiment classification and aspect category detection. The model then generates pseudo-labels for Korean data, which are used for training the final model. This final model is trained either on a combination of the translated data and the pseudo-labelled data or solely on the pseudo-labelled data. To enhance the quality of the pseudo-labels, dual filtering techniques are employed, including LaBSE-based filtering and confidence score filtering. LaBSE, originally developed as a dual-encoder for machine translation, proves equally effective for monolingual tasks such as semantic textual similarity (STS) and generating high-quality training data. This cost-efficient approach demonstrates how resource gaps can be bridged, providing an effective solution for ABSA in low-resource languages.

5.3. Aspect Category Detection

Aspect category detection involves identifying the relevant aspect categories discussed in a given sentence, which typically belong to a predefined set of categories that are often domain-specific. For instance, given the sentence “*The tea is great, but the ambience is disappointing*”, an ACD method should accurately predict the categories of “*drinks*” and “*ambience*”.

Compared to the ATE task, ACD offers advantages from two main perspectives. First, while ATE focuses on predicting individual aspect terms, predicted categories of ACD represent an aggregated outcome, providing a more concise summary of the opinion targets. Second, ACD can identify opinion targets even when they are not explicitly mentioned. For example, in the sentence “*The wait was long, and the lunch was overpriced*”, ACD can recognize two aspect categories: “*service*” and “*price*”, whereas ATE would not be applicable in this scenario.

The ACD task is often framed as a multi-label sequence-level classification problem, where each category is a label. Phan et al. [27] employ mBERT and XLM-R in English, Spanish, French, Dutch, Russian, and Turkish, demonstrating the superior results with XLM-R. Similarly, Dang Van Thin and Nguyen [29] explore the same models and languages except for Turkish, training the models on different combinations of source languages. Additionally, Nam [30] proposes a pseudo-labelling framework for low-resource languages, which leverages multilingual BERT fine-tuned on translated data to generate high-quality pseudo-labels, as detailed in the ASC task.

5.4. End-to-End ABSA

The end-to-end ABSA task aims to simultaneously identify aspect terms and their corresponding sentiment polarities within a given sentence. This task is unique in cross-lingual settings as it is the only compound task explored extensively. The focus on end-to-end ABSA has grown significantly with the advent of multilingual pre-trained language models.

The E2E-ABSA task is often tackled using a unified tagging scheme. The first part of the scheme, such as BIO or BIOES, marks the boundaries of aspect terms, while the second part denotes the sentiment polarity. For example, the BIOES scheme uses B for the beginning, I for inside, O for outside, E for the end, and S for singleton, denoting the boundaries of aspect terms, while the sentiment polarity is indicated as either positive (POS), negative (NEG), or neutral (NEU). For instance, B-NEU refers to the beginning of an aspect term with a neutral sentiment polarity. This tagging structure allows end-to-end ABSA to be modelled as a token classification problem using a sequence tagger. Table 5 provides an example of the BIOES and BIO tagging formats.

Sattar et al. [31] use part-of-speech (POS) tagging to extract tokens from the text, pre-process them by tasks like removing punctuation, and then map the tokens between languages using bilingual dictionaries. They utilize mBERT to generate embeddings and feed both the POS tags and embeddings into an attention-based CNN for further processing.

Szołomicka and Kocon [32] present TrAsp, a multilingual and language-agnostic approach to ABSA that employs Transformer-based methods. TrAsp utilizes a Transformer-based embedding model with two linear layers and one dropout layer, achieving superior performance compared to existing techniques. The method is evaluated on the MultiAspectEmo dataset (covering Polish and its machine-translated versions in English, Czech, Spanish, French, and Dutch) and the SemEval-2016 dataset. The results demonstrate that multilingual pre-trained models like XLM-R significantly outperform language-agnostic models such as LaBSE. The authors also compare the cross-lingual capabilities of LaBSE and XLM-R, highlighting a substantial drop in cross-lingual performance relative to monolingual settings, with XLM-R consistently outperforming LaBSE. The study underscores the advantages of multilingual training for cross-lingual knowledge transfer, achieving higher F1 scores, and investigates the potential of compressed models for resource-efficient ABSA. The multilingual capabilities of pre-trained models are shown to be pivotal in enabling robust cross-lingual performance.

Wu et al. [33] evaluate zero-shot multilingual ABSA with LLMs for the E2E-ABSA task across English, French, Spanish, Dutch, and Russian using the SemEval-2016 dataset. They explore various prompting strategies, including vanilla zero-shot, chain-of-thought, self-improvement, self-debate, and self-consistency. Their findings reveal that while LLMs show potential for multilingual ABSA, they generally underperform compared to fine-tuned, task-specific models. Performance varies significantly by language, with higher-resource languages like English achieving better results than low-resource languages and simpler zero-shot prompts often outperforming more complex strategies in these languages. However, LLMs face challenges in extracting fine-grained structured information and struggle with reasoning tasks, especially in multi-turn dialogue contexts. Closed-source models generally outperform open-source ones, and few-shot learning improves performance, although context length remains a constraint. Furthermore, cross-lingual evaluation, where models fine-tuned on English data are tested on other languages, shows significant performance drops in morphologically different languages such as Russian. These findings highlight the potential of LLMs for multilingual ABSA while emphasizing the need for further research to address their dependencies on task, language, and configuration, as well as to optimize their application for cross-lingual tasks.

Li et al. [11] explore their span-to-span translation mapping strategy, described in Section 4.2, by training XLM-R and mBERT in three different settings: using only the source language data (ZERO-SHOT), using only translated target language data (TRANSLATION-TA), and using a combination of both (BILINGUAL-TA). While their approach proved effective for semantic role labelling, they find that for cross-lingual ABSA, fine-tuning models on the original source language data consistently outperformed training on translated data or combinations of both, which they attribute to the model's lower tolerance to translation errors. To address this, they proposed a model parameter warm-up on subsets of translated data from multiple target languages (MTL-WS) outperforming the multilingual training baseline (MTL-TA) with models fine-tuned on a combination of data translated to multiple target languages.

Zhang et al. [12] demonstrate the efficacy of their alignment-free translation mechanism discussed in Section 4.2 in several configurations: training on pseudo-labelled target language data alone (TRANSLATION-AF), combining translated and source data (BILINGUAL-AF), and training on both original and translated data alongside aspect-code-switched data, where aspect terms are swapped between languages (ACS). In this setting, they initialize a student model using parameters from a model pre-trained on the translated target language data. The student model is then fine-tuned to match the soft probability distribution predicted by the teacher model for each token, effectively learning from the teacher's output rather than the original hard labels. This approach helps the student model generalize better to the target language by approximating the teacher's predictions, allowing it to capture more nuanced sentiment information. They experimented with single-teacher distillation, using a teacher trained on the ACS dataset (ACS-DISTIL-S), and multi-teacher distillation, where each teacher is trained on different datasets, such as translated and source data or translated and aspect-code switched data (ACS-DISTIL-M). Furthermore, they experiment with multilingual settings, utilizing multilingual data with alignment-free label projection (MTL-AF) and aspect-code-switched data (MTL-ACS), and distillation on multilingual unlabelled data (MTL-ACS-D).

Lin et al. [13, 14] further utilize the ACS data and add some additional improvements. They employ contrastive learning to minimize the distance between samples with the same label in different semantic spaces [13]. Specifically, they use token-level contrastive learning to align token embeddings with similar labels (CL-XABSA (TL)) and sentiment-level contrastive learning for samples at the sentiment level (CL-XABSA (SL)). Additionally, they address class imbalances with dynamic weighted loss, which shifts the model's attention towards underperforming sentiment categories, and introduce anti-decoupling techniques to enhance semantic information utilization (EQUI-ABSA) [14].

Table 6

Cross-lingual end-to-end ABSA results of different methods with English as the source language and other languages as target ones compared to supervised settings (SUPERVISED). The best results for each model and language are in **bold**. *The average results exclude Turkish.

Source	Method	mBERT						XLM-R					
		Es	Fr	Nl	Ru	Tr	Avg*	Es	Fr	Nl	Ru	Tr	Avg*
[12]	SUPERVISED	67.88	61.80	56.80	58.87	–	61.34	71.93	67.44	64.28	64.93	–	67.15
	ZERO-SHOT	57.32	45.60	42.68	36.01	26.58	45.40	67.10	56.43	59.03	56.80	46.21	59.84
[11]	TRANSLATION-TA	50.74	40.76	47.13	41.67	22.04	45.08	58.10	47.00	56.19	50.34	40.24	52.91
	BILINGUAL-TA	51.23	41.00	49.72	43.67	22.64	46.41	61.87	49.34	58.64	52.89	41.44	55.69
[12]	TRANSLATION-AF	59.74	48.03	49.73	50.17	–	51.92	66.61	57.07	61.26	59.55	–	61.12
	BILINGUAL-AF	60.23	48.05	49.83	51.24	–	52.34	68.04	57.91	60.80	60.81	–	61.89
[12]	ACS	59.99	49.65	51.19	52.09	–	53.23	67.32	59.39	62.83	60.81	–	62.59
	ACS-DISTIL-S	62.04	52.23	52.72	53.00	–	55.00	68.93	61.00	62.89	60.97	–	63.45
	ACS-DISTIL-M	62.91	52.25	53.40	54.58	–	55.79	69.24	59.90	63.74	62.02	–	63.73
[13]	CL-XABSA(TL)	60.64	48.53	50.96	50.77	–	52.73	–	–	–	–	–	–
	CL-XABSA(SL)	61.62	49.50	50.64	50.65	–	53.10	–	–	–	–	–	–
[14]	EQUI-XABSA	63.08	50.08	51.85	52.59	–	54.40	69.56	60.68	61.31	62.34	–	63.47

Table 7

Multilingual end-to-end ABSA results of different methods. The best results for each model and target language are in **bold**. *The average results exclude Turkish.

Source	Method	mBERT						XLM-R					
		Es	Fr	Nl	Ru	Tr	Avg*	Es	Fr	Nl	Ru	Tr	Avg*
[11]	MTL-TA	54.14	40.72	49.06	43.89	25.77	46.95	63.56	52.80	60.37	55.67	43.04	58.10
	MTL-WS	58.18	46.93	49.87	44.88	29.78	49.96	68.60	57.96	61.24	59.74	45.58	61.89
[12]	MTL-AF	59.31	50.00	53.16	50.04	–	53.13	68.20	59.68	63.33	61.02	–	63.06
	MTL-ACS	59.59	50.74	53.33	51.61	–	53.82	68.88	59.19	63.06	61.92	–	63.26
	MTL-ACS-D	62.05	53.56	53.56	53.87	–	55.76	70.38	62.17	65.98	62.79	–	65.33

Table 6 presents the cross-lingual results with mBERT and XLM-R on the SemEval-2016 restaurant dataset, with English as the source language and other languages as targets. The best results are typically achieved with ACS-DISTIL-M [12], which leverages aspect-code switching and distillation on unlabelled target language data, and EQUI-ABSA [14], which addresses class imbalances. Table 7 shows the multilingual result where MTL-ACS-D [12] achieves the best results by combining aspect code-switching and multilingual distillation on unlabelled target language data.

6. Related Research

In this section, we describe some of the related research for ABSA, focusing on innovative methods in recent years, languages other than English, multilingual ABSA, cross-lingual sentiment analysis, and LLMs for ABSA. Additionally, we shortly summarize advancements in multilingual and cross-lingual sentiment analysis. Finally, we discuss the possibilities of utilizing large language models for ABSA.

6.1. Monolingual ABSA

Ma et al. [81] address the TASD task with a neural architecture combining a bidirectional LSTM sequence encoder and a hierarchical attention mechanism. Their model employs target-level attention to highlight sentiment-relevant parts of aspect terms and sentence-level attention to identify aspect evidence across the sentence. Additionally, they propose

Sentic LSTM, an enhanced LSTM cell incorporating affective commonsense knowledge. Liang et al. [82] introduce Sentic GCN, a graph convolutional network that augments sentence dependency graphs with affective knowledge from SenticNet. This model captures both contextual and affective relationships between words and aspects by integrating affective dependencies and focusing on aspect-specific sentiment. Both approaches can be extended to cross-lingual settings by incorporating cross-lingual word embeddings or machine translation.

D’Aniello et al. [83] present KnowMIS-ABSA, a framework emphasizing distinctions among key concepts such as sentiment, emotion, affect, and opinion. Unlike traditional methods, it advocates for tailored metrics and tools to measure these dimensions separately, ensuring a more nuanced understanding of user opinions. A qualitative case study demonstrates its advantages, particularly in marketing, where differentiating sentiments from affective reactions can enhance strategies.

Recent English ABSA research predominantly adopts the sequence-to-sequence modelling paradigm due to its flexibility in handling compound tasks involving three or four sentiment elements. Annotation- and extraction-style modelling paradigms by Zhang et al. [44] showcase the viability of a generative framework for ABSA. Later work [44] introduces a novel approach to the ASQP task, predicting sentiment quads in natural language format, facilitating more intuitive sentiment analysis results. Gao et al. [84] propose a comprehensive framework that addresses multiple ABSA tasks by integrating element prompts, while Mao et al. [61] generate tuples as tree paths and filter valid ones. Most of these methods generate sentiment elements in a fixed order, which may not be optimal. Hu et al. [85] investigate the impact of sentiment element ordering on ABSA outcomes, optimizing the sequence generation process. Building on this, Gou et al. [62] utilize multiple-ordered permutations to enhance sentiment element prediction, merging results from different orders at the cost of extended training times. Xianlong et al. [63] combine sequence-to-sequence and token classification paradigms. The encoder labels aspect and opinion terms, merging results with the model’s text output. All these methods can be adapted to cross-lingual transfer using multilingual Transformer-based models, optionally enhanced by machine translation.

For languages beyond English, Czech ABSA research has gained traction. Earlier studies [45, 46, 86] focus on simple tasks with two sentiment elements, combining carefully selected features with classifiers like CRF and logistic regression. These methods are adaptable to cross-lingual settings via machine translation. Later study [87] explores convolutional and recurrent neural networks for ACS and ACD tasks, adaptable using machine translation or cross-lingual embeddings. Recent studies leverage Transformer-based models. For instance, Šmíd and Přibáň [64] introduce prompt-based methods for Czech ABSA, demonstrating the effectiveness of multilingual sequence-to-sequence and Czech-specific BERT-base models and in-domain pre-training in few-shot settings. Similarly, another work [88] enhances ABSA with semantic role labelling in a multitask learning framework, improving ACD and ASC tasks in both Czech and English. Šmíd et al. [47] deliver state-of-the-art Czech ABSA results using Transformer-based methods with both multilingual and Czech-specific models. The modern approaches are adaptable to cross-lingual settings through multilingual models and machine translation.

6.1.1. Relationship to Cross-lingual ABSA

The approaches discussed can be adapted to cross-lingual ABSA using cross-lingual transfer techniques from Section 4, such as cross-lingual word embeddings, machine translation, multilingual pre-trained models, or a combination of them. A key distinction between monolingual and cross-lingual research lies in their focus: monolingual research aims to improve ABSA task performance, while cross-lingual research prioritizes enhancing transfer capabilities across languages. Combining the strengths of both fields could significantly improve cross-lingual ABSA. For instance, leveraging multilingual models with machine translation and target-language distillation alongside high-performing monolingual approaches like tagging-assisted generation could yield substantial advancements.

6.2. Multilingual and Few-Shot Cross-Domain ABSA

García-Pablos et al. [89] present W2VLDA, an almost unsupervised system for ABSA that requires only minimal supervision, addressing the limitations of resource-intensive supervised approaches. W2VLDA combines latent Dirichlet allocation (LDA) [90], continuous word embeddings, and a maximum entropy classifier to classify domain-specific aspects and sentiment polarity. It operates on unlabelled textual corpora, requiring only a single seed word per domain aspect and two polarity seed words (positive and negative). The system generates weighted lists of aspect terms, opinion words, and classified sentences, aligning these outputs with predefined domain aspects. Evaluated on the SemEval-2016 5 dataset across multiple domains (restaurants, hotels, electronic devices) and languages (English, Spanish, French, Dutch), W2VLDA achieves competitive results. Its reliance on minimal annotated data and automatic

separation of aspect terms and opinion words makes it adaptable to diverse domains and languages, offering an alternative to cross-lingual methods for ABSA.

He et al. [91] introduce LGCF, a multilingual framework for ABSA that effectively captures local and global context features to predict sentiment polarity. LGCF uses BERT-based embeddings with distinct layers to model local and global contexts. Local context modelling is enhanced by multi-headed self-attention (MHSA), dynamic masking (CDM), and dynamic weighting (CDW). A combination of bidirectional GRU [92] and CNN is employed for global context modelling, enabling robust feature learning for aspect-sentiment correlations. Validated on Chinese and English datasets, LGCF achieves state-of-the-art performance, demonstrating the advantages of combined local and global context modelling.

Rana et al. [93] propose a hybrid model for multilingual sentiment analysis that combines BERT with a lexicon-based approach to improve polarity estimation for reviews and tweets. The model incorporates multilingual conversion to translate non-English content into English, effectively handling multilingual text. Processing involves stages like preprocessing, aspect extraction, aspect engineering, and polarity calculation, leveraging resources such as SentiWordNet and part-of-speech taggers. While initially focused on multilingual scenarios, this approach can be adapted for cross-lingual settings by training in one language and testing in another, as machine translation is already an integral part of the workflow.

Another study [94] explores cross-domain few-shot classification for ABSA, where models trained in one domain (e.g. hotel reviews) are evaluated in another (e.g. restaurant reviews) with minimal labelled target data. While primarily focused on domain adaptation within a single language, the methods bear similarities to cross-lingual ABSA, where knowledge is transferred across languages instead of domains. Using the Polish AspectEmo dataset, the study compares gradient-based learning, zero-shot approaches, and few-shot methods like ProtoNet [95] and NNShot [96]. Few-shot methods outperform other approaches, with NNShot offering a slight edge. These methods leverage embeddings to predict labels in the target domain by comparing support and query set similarities. The findings suggest the potential for few-shot learning in cross-lingual ABSA, where models could benefit from limited examples in the target language.

6.2.1. Relationship to Cross-lingual ABSA

Multilingual ABSA research often overlaps with cross-lingual ABSA techniques, such as machine translation and multilingual pre-trained models, making adaptation to cross-lingual settings straightforward. However, key differences must be addressed. Multilingual approaches typically include the target language in the training data, allowing models to learn language-specific nuances during training. In contrast, zero-shot cross-lingual transfer relies solely on source language data, requiring the model to generalize across languages.

Using multiple source languages can enhance performance by providing diverse linguistic patterns but may also hinder results if poorly chosen [97]. Strategies like machine translation to convert source language data into the target language or carefully selecting source languages can mitigate these challenges. Machine translation-based approaches can augment original datasets, either supplementing or replacing them, to provide better language-specific adaptation. Few-shot cross-lingual approaches could also leverage minimal labelled data in the target language to fine-tune models trained on source language data, combining the strengths of multilingual and cross-lingual methods.

6.3. Multilingual and Cross-lingual Sentiment Analysis

This section provides an overview of multilingual and cross-lingual sentiment analysis approaches, focusing on document-level sentiment classification. Unlike aspect-based sentiment analysis, which requires fine-grained analysis at the aspect level, document-level sentiment analysis assigns sentiment to an entire document, making it a comparatively simpler task. Additionally, it is a more examined task than cross-lingual ABSA.

Early approaches to cross-lingual sentiment analysis frequently leverage machine translation [98, 99, 100, 101, 102], often translating English training data into other languages to train sentiment models. Other work [103] utilizes translation for specific resources, such as lexical sentiment dictionaries like SentiWordNet, to enrich datasets and improve sentiment analysis in target languages.

Another significant line of research focuses on developing cross-lingual word embeddings specifically tailored for sentiment analysis [104]. These embeddings are employed with models like SVM and neural networks such as CNN and LSTM [105, 106]. Researchers also introduce approaches to align word embeddings across languages using small bilingual dictionaries or sentiment information derived from the source language, allowing for efficient transfer of sentiment knowledge to target languages [107].

In recent years, Transformer-based architectures have dominated cross-lingual sentiment analysis. Multilingual versions of BERT and related models have proven particularly effective [106, 108, 109, 110]. Studies have further enhanced these models with data augmentation techniques, such as augmenting datasets through machine translation [111]. Some studies [112, 113, 114, 115] further investigate hybrid strategies with multilingual pre-trained models and machine translation, such as mixing words between languages to enhance training, and focus on low-resource languages that inherently include code-mixed data, such as anglicisms commonly found in social media texts.

Further advancements include the exploration of zero-shot and few-shot learning strategies for cross-lingual classification tasks, enabling sentiment analysis in languages unseen during pre-training [116]. Researchers have also evaluated the effectiveness of linear transformations for zero-shot sentiment classification and subjectivity classification tasks, providing alternatives to Transformer-based methods [105, 106].

6.3.1. Relationship to Cross-lingual ABSA

Compared to multilingual and cross-lingual sentiment analysis, cross-lingual ABSA employs many of the novel and best-performing approaches for cross-lingual transfer, such as modern pre-trained multilingual language models, machine translation, mixing words between languages, and cross-lingual embeddings. However, there is limited scope for further direct adoption of techniques from cross-lingual sentiment analysis into cross-lingual ABSA, as many of these methods have already been utilized.

Cross-lingual ABSA introduces additional complexities that are not as prominent in document-level sentiment analysis. One such challenge is the precise alignment of aspect and opinion term labels, which is crucial for accurate aspect-based sentiment predictions. This alignment task is inherently difficult due to the nuanced relationship between aspect terms and their corresponding sentiment expressions, especially when crossing linguistic boundaries. Furthermore, cross-lingual ABSA often requires greater attention to language-specific features, such as idiomatic expressions, colloquialisms, or slang, which may vary significantly across languages. These nuances demand a more tailored approach and further complicate the transfer of knowledge from one language to another.

6.4. Large Language Models for ABSA

Large language models refer to Transformer-based architectures with parameter counts exceeding 1 billion or even 10 billion, depending on the source [117, 65]. These generative models take text as input and produce text as output, with most of them based on the decoder stack of the Transformer. Examples of LLMs include PaLM [118], GPT-4 [119], and LLaMA [120, 121, 122]. LLMs are pre-trained on massive text corpora, and many are further tuned to follow instructions. This tuning has shifted the paradigm from traditional fine-tuning to *prompting*, where tasks are defined directly in the input prompt without requiring labelled training data. Prompting can be zero-shot (task definition only) or few-shot (task definition with examples, also called in-context prompting), enabling effective solutions in scenarios with limited labelled data. This capability is especially beneficial in cross-lingual settings, where annotated data for specific tasks and domains might not exist in all target languages.

While LLMs have demonstrated state-of-the-art performance across many NLP tasks, including sentiment analysis [65, 106, 123], they often underperform compared to smaller, fine-tuned models for ABSA tasks [62, 65]. However, fine-tuning LLMs specifically for ABSA has achieved impressive results, including end-to-end ABSA in English [124] and complex triplet and quadruplet tasks [66]. Early study [33] focusing on the multilingual performance of LLMs has broadly explored the cross-lingual abilities of fine-tuned LLMs with varied results.

The performance of LLMs in ABSA tasks can be affected by inconsistencies in dataset annotations, particularly for opinion terms. Some datasets label modifiers like “very” [34], while others omit them [40, 42, 43], leading to inconsistencies that affect evaluation metrics. Fine-tuned models can adapt to such patterns if the annotations are consistent. However, LLMs operating in zero-shot or few-shot settings may produce outputs that do not align precisely with the evaluation criteria. To address these challenges, some better evaluation techniques of LLMs for ABSA might be needed. Additionally, techniques such as *chain-of-thought* prompting [125] and reasoning-based multi-prompt methods [126] may improve LLM performance in complex ABSA tasks.

Recent studies illustrate the nuanced performance of LLMs in ABSA. Filip et al. [127] use several LLMs with smaller BERT-like models for the ASC task in Czech, Slovakian, Polish, and Hungarian. Using a dataset of tweets collected via the Twitter/X academic API in 2023 and manually annotated for sentiment towards Russia and Ukraine, they fine-tuned models such as LLaMA 2 [121], LLaMA 3 [122], Mistral [128], BERT, and BERTweet, with GPT-4 as a reference. LLaMA 2 and Mistral achieved state-of-the-art results, while LLaMA 3 and BERTweet performed worse, likely due to the limited tunability of LLaMA 3 and outdated pre-training data of BERTweet. Fine-tuning proved

significantly more effective than in-context learning, particularly for multilingual and culturally complex datasets. Translating data into English improved performance across all models; however, fine-tuned models like LLaMA 2 exhibited strong adaptability even in the original languages, surprising given their English-centric training. Challenges included misclassification in Polish tweets due to cultural nuances and limitations such as model rigidity and smaller pre-trained parameter sets. The study highlighted the effectiveness of fine-tuning on small datasets and the importance of task-specific and model-specific adaptations.

Mughal et al. [129] examine older deep neural networks, including LSTMs and BERT-based models, alongside LLMs for ATE and ACD tasks on several English datasets. Their findings emphasized PaLM's robust performance, often surpassing BERT-like models across diverse domains, while identifying its limitations on challenging datasets like MAMS, which contain multiple aspects and sentiments in a single review. GPT-3.5 delivered competitive MAMS results comparable to specialized models like ATAE-LSTM, showcasing LLMs' ability to handle complex datasets with minimal fine-tuning. The study demonstrated LLMs' strengths in transfer learning, contextual reasoning, and domain versatility, exposing their gaps in handling nuanced tasks. Older models like ATAE-LSTM and DeBERTa [130] excelled in domain-specific scenarios, underscoring the continued relevance of specialized training. The authors advocate exploring advanced LLMs, such as GPT-4, and refining models to better address nuanced sentiment and cross-domain challenges, offering valuable insights into the evolving ABSA landscape.

Luo et al. [37] introduce a groundbreaking framework for multimodal conversational ABSA, addressing gaps in integrating multimodality, conversational context, fine-grained sentiment granularity, dynamic sentiment shifts, and cognitive causal reasoning. Two new tasks were proposed: panoptic sentiment sextuple extraction, which identifies holder, target, aspect, opinion, sentiment, and rationale from multi-turn multimodal dialogues; and sentiment flipping analysis, which detects dynamic sentiment transformations and their causal factors. The PanoSent dataset, featuring multilingual, multimodal, and multi-scenario annotations, was created to benchmark these tasks. To tackle the challenges, a chain-of-sentiment reasoning framework, a novel multimodal LLM (Sentica), and a paraphrase-based verification mechanism were devised, outperforming strong baselines. This study's multilingual approach (including English, Chinese, and Spanish) and multimodal emphasis offer a robust foundation for advancing cross-lingual ABSA, addressing gaps created by the lack of annotated data for complex tasks.

Furthermore, LLMs show potential for data augmentation [131], with studies leveraging LLMs to generate additional training data [132, 133], including for English ABSA [134]. Such approaches could generate augmented data in target languages; however, their effectiveness may heavily depend on the quality of the generated data.

Despite their advantages, LLMs face several challenges. Closed-source models like GPT-4 are expensive to use and may raise privacy concerns for real-world applications. Open-source alternatives like LLaMA require substantial hardware resources for fine-tuning. For instance, fine-tuning a 13B-parameter model in full precision requires over 200 GB of GPU memory. Techniques like QLoRA [135], which employ a 4-bit quantized frozen backbone with learnable LoRA [136] parameters, reduce memory requirements and enable fine-tuning on consumer GPUs. However, even with these optimizations, LLMs remain resource-intensive compared to smaller, specialized models.

A significant limitation of many LLMs is their English-centric design [121, 137]. While newer multilingual LLMs like LLaMA 3.1 [122] and Aya [138, 139] have emerged, they often support fewer languages than multilingual pre-trained models like XLM-R. Limited language support results from the high cost and effort required to obtain quality multilingual data for instruction tuning, with quality being more critical than quantity [140].

One major advantage of LLMs is their ability to generate results without labelled training data, which is costly and time-consuming to produce. This makes them especially appealing for tasks in under-resourced languages or domains where annotated datasets are scarce.

6.5. Relationships to Cross-lingual ABSA

Large language models have demonstrated effectiveness in simpler ABSA tasks under zero-shot and few-shot settings, offering quick solutions when labelled data is unavailable. However, smaller fine-tuned models often outperform LLMs for more complex tasks, such as predicting multiple sentiment element tuples or handling nuanced cross-lingual datasets. Techniques like chain-of-thought prompting and task-specific augmentations can enhance LLM performance in these scenarios. LLMs are also valuable for bridging language gaps, particularly when annotated data is available in a source language but not in the target language. Despite limitations such as high hardware requirements and performance gaps on challenging ABSA tasks, LLMs are useful for rapid prototyping, data augmentation, and addressing under-resourced languages.

Compared to traditional cross-lingual ABSA approaches, multilingual LLMs offer a notable advantage by eliminating the need for annotated data, even in a language different from the target language. LLMs can achieve acceptable performance for some tasks using zero-shot and few-shot prompts. However, research consistently shows that fine-tuned models outperform LLMs in most cases. That said, as LLMs continue to evolve, future iterations may match or exceed the performance of smaller multilingual models specifically fine-tuned for ABSA tasks.

7. Challenges and Future Research

While monolingual ABSA has seen considerable progress with new approaches and tasks, cross-lingual ABSA remains significantly more challenging and underexplored. This section highlights key challenges and potential directions for future research.

7.1. Quests for More Diverse Multilingual Datasets

As noted in Section 2.3, the number of available multilingual ABSA datasets for cross-lingual transfer is extremely limited. Although some datasets cover multiple domains, only the SemEval-2016 dataset provides data in more than two languages. Most datasets focus on aspect terms and sentiment polarity, with only one offering aspect category annotations and none including opinion term annotations. This lack of annotated sentiment elements restricts the range of tasks that can be studied. Additionally, many datasets contain relatively few sentences, which hinders reliable comparison of models, especially for large pre-trained models with millions of parameters. To advance cross-lingual ABSA research, there is a need for more diverse datasets with annotations for multiple sentiment elements across different languages.

7.2. Cross-lingual ABSA Tasks

As discussed in Section 2.2, while numerous ABSA tasks exist, only a few have been explored in cross-lingual settings. This limitation is partly due to the lack of annotated sentiment elements, particularly opinion terms, which restricts the ability to explore opinion-related tasks. Although the SemEval-2016 dataset supports tasks with three linked sentiment elements (i.e. the T ASD task), research has largely focused on tuple-based tasks, specifically E2E-ABSA, and single tasks focusing only on one sentiment element. Expanding the exploration of cross-lingual ABSA tasks requires more comprehensive multilingual datasets.

7.3. Sequence-to-Sequence Modelling Paradigm and LLMs in Cross-lingual ABSA

The sequence-to-sequence modelling paradigm, as discussed in Section 3.3, has recently gained significant traction in English ABSA research [44, 60, 61, 62, 63]. However, its application in cross-lingual settings remains largely unexplored [141]. This approach, particularly suitable for tasks involving multiple sentiment elements, holds great potential for addressing the complexities of cross-lingual ABSA.

Fine-tuned LLMs have demonstrated remarkable performance in monolingual ABSA [66]. However, their use in cross-lingual contexts has seen only limited investigation, except for a minor study focused on multilingualism [33]. The development of multilingual large language models [122] presents an exciting opportunity for advancing cross-lingual ABSA. Fine-tuning these models could significantly improve addressing linguistic diversity and task complexity.

Looking ahead, the continual advancement in LLM performance for ABSA tasks opens the door to novel approaches that bypass fine-tuning entirely. Instead, LLMs could be employed with zero-shot or few-shot prompting strategies. Such methods eliminate the need for annotated training datasets, addressing the challenges posed by the scarcity of labelled resources in many languages. This paradigm shift could enable broader accessibility and applicability of cross-lingual ABSA solutions.

7.4. Adopting Modern Approaches from Monolingual ABSA

Cross-lingual ABSA research has predominantly concentrated on enhancing language transfer techniques to bridge the gap between source and target languages. Conversely, monolingual ABSA research has prioritized refining task-specific methodologies, resulting in several significant advancements tailored to ABSA. Despite their demonstrated success, many of these innovations from monolingual ABSA research have yet to be fully incorporated into cross-lingual studies.

Integrating state-of-the-art methods from monolingual ABSA, such as tagging-assisted generation or multi-view prompting, with the best-performing cross-lingual transfer techniques from cross-lingual ABSA could yield

substantial performance improvements. Bridging these two lines of research offers a promising direction for future studies. By leveraging the complementary strengths of monolingual task-specific innovations and cross-lingual transfer methodologies, researchers can push the boundaries of what is achievable in cross-lingual ABSA.

7.5. Dependency on Machine Translation

Although multilingual pre-trained models offer inherent cross-lingual capabilities, they are often combined with machine translation to improve performance [12, 13, 14]. However, studies have shown that machine translation can sometimes degrade performance due to translation quality and label misalignment [11]. Recent advancements in LLMs provide new possibilities for cross-lingual ABSA, as LLM-based data augmentation has emerged as a promising technique to enhance performance [142], with some studies using LLMs to generate additional data [132, 133, 143]. Future research could explore replacing reliance on machine translation with LLM-generated data.

8. Conclusion

This survey comprehensively reviews cross-lingual aspect-based sentiment analysis, representing the first dedicated survey to focus exclusively on this emerging field. It begins by outlining the necessary background, including the definition of aspect-based sentiment analysis, the four sentiment elements, and the various tasks and datasets used in cross-lingual ABSA research. Next, it describes the different modelling paradigms employed in ABSA, followed by an overview of approaches to cross-lingual transfer that enable sentiment analysis across languages, with examples of how these techniques have been applied in recent studies. We also offer a detailed summary of the ABSA tasks explored in cross-lingual settings, highlighting the methods used to address each task. The review also highlights the relevance of monolingual and multilingual ABSA research, as well as ABSA approaches with LLMs, in addressing challenges in cross-lingual ABSA. Finally, we discuss the main challenges facing the field and suggest potential directions for future research to advance cross-lingual ABSA.

CRedit authorship contribution statement

Jakub Šmíd: Conceptualization, Validation, Formal analysis, Investigation, Resources, Writing – original draft, Visualization, Supervision, Project administration. **Pavel Král:** Writing – review & editing, Supervision.

Acknowledgements

This work has been partly supported by the OP JAC project “R&D of Technologies for Advanced Digitization in the Pilsen Metropolitan Area (DigiTech)” No.: CZ.02.01.01/00/23_021/0008436 co-financed by the European Union. 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.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work the authors used the ChatGPT¹ tool in order to readability, improve language, and correct grammatical errors. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication. The tool was not used to analyse data or derive insights during the research process.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] B. Liu, Sentiment analysis and subjectivity., Handbook of natural language processing 2 (2010) 627–666.

¹<https://chat.openai.com>

- [2] W. Zhang, X. Li, Y. Deng, L. Bing, W. Lam, A survey on aspect-based sentiment analysis: Tasks, methods, and challenges, *IEEE Transactions on Knowledge and Data Engineering* (2022).
- [3] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. M. Jiménez-Zafra, G. Eryigit, SemEval-2016 task 5: Aspect based sentiment analysis, in: S. Bethard, M. Carpuat, D. Cer, D. Jurgens, P. Nakov, T. Zesch (Eds.), *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, Association for Computational Linguistics, San Diego, California, 2016, pp. 19–30. URL: <https://aclanthology.org/S16-1002>. doi:10.18653/v1/S16-1002.
- [4] P. Lambert, Aspect-level cross-lingual sentiment classification with constrained SMT, in: C. Zong, M. Strube (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)*, Association for Computational Linguistics, Beijing, China, 2015, pp. 781–787. URL: <https://aclanthology.org/P15-2128>. doi:10.3115/v1/P15-2128.
- [5] X. Zhou, X. Wan, J. Xiao, Clopinionminer: Opinion target extraction in a cross-language scenario, *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 23 (2015) 619–630.
- [6] J. Barnes, P. Lambert, T. Badia, Exploring distributional representations and machine translation for aspect-based cross-lingual sentiment classification, in: Y. Matsumoto, R. Prasad (Eds.), *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, The COLING 2016 Organizing Committee, Osaka, Japan, 2016, pp. 1613–1623. URL: <https://aclanthology.org/C16-1152>.
- [7] M. S. Akhtar, P. Sawant, S. Sen, A. Ekbal, P. Bhattacharyya, Solving data sparsity for aspect based sentiment analysis using cross-linguality and multi-linguality, in: M. Walker, H. Ji, A. Stent (Eds.), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, Association for Computational Linguistics, New Orleans, Louisiana, 2018, pp. 572–582. URL: <https://aclanthology.org/N18-1053>. doi:10.18653/v1/N18-1053.
- [8] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: J. Burstein, C. Doran, T. Solorio (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)*, Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: <https://aclanthology.org/N19-1423>. doi:10.18653/v1/N19-1423.
- [9] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: D. Jurafsky, J. Chai, N. Schluter, J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Online, 2020, pp. 8440–8451. URL: <https://aclanthology.org/2020.acl-main.747>. doi:10.18653/v1/2020.acl-main.747.
- [10] J. Hu, S. Ruder, A. Siddhant, G. Neubig, O. Firat, M. Johnson, XTREME: A massively multilingual multi-task benchmark for evaluating cross-lingual generalisation, in: H. D. III, A. Singh (Eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, PMLR, 2020, pp. 4411–4421. URL: <https://proceedings.mlr.press/v119/hu20b.html>.
- [11] X. Li, L. Bing, W. Zhang, Z. Li, W. Lam, Unsupervised cross-lingual adaptation for sequence tagging and beyond, *arXiv preprint arXiv:2010.12405* (2020).
- [12] W. Zhang, R. He, H. Peng, L. Bing, W. Lam, Cross-lingual aspect-based sentiment analysis with aspect term code-switching, in: M.-F. Moens, X. Huang, L. Specia, S. W.-t. Yih (Eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 2021, pp. 9220–9230. URL: <https://aclanthology.org/2021.emnlp-main.727>. doi:10.18653/v1/2021.emnlp-main.727.
- [13] N. Lin, Y. Fu, X. Lin, D. Zhou, A. Yang, S. Jiang, Cl-xabsa: Contrastive learning for cross-lingual aspect-based sentiment analysis, *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2023).
- [14] N. Lin, M. Zeng, X. Liao, W. Liu, A. Yang, D. Zhou, Addressing class-imbalance challenges in cross-lingual aspect-based sentiment analysis: Dynamic weighted loss and anti-decoupling, *Expert Systems with Applications* 257 (2024) 125059.
- [15] S. Poria, D. Hazarika, N. Majumder, R. Mihalcea, Beneath the Tip of the Iceberg: Current Challenges and New Directions in Sentiment Analysis Research, *IEEE Transactions on Affective Computing* 14 (2023) 108–132.
- [16] K. Schouten, F. Frasnar, Survey on aspect-level sentiment analysis, *IEEE Transactions on Knowledge and Data Engineering* 28 (2016) 813–830.
- [17] A. Nazir, Y. Rao, L. Wu, L. Sun, Issues and challenges of aspect-based sentiment analysis: A comprehensive survey, *IEEE Transactions on Affective Computing* 13 (2022) 845–863.
- [18] G. Brauwiers, F. Frasnar, A survey on aspect-based sentiment classification, *ACM Comput. Surv.* 55 (2022).
- [19] G. S. Chauhan, R. Nahta, Y. K. Meena, D. Gopalani, Aspect based sentiment analysis using deep learning approaches: A survey, *Computer Science Review* 49 (2023) 100576.
- [20] T. Zhao, L. ang Meng, D. Song, Multimodal aspect-based sentiment analysis: A survey of tasks, methods, challenges and future directions, *Information Fusion* 112 (2024) 102552.
- [21] B. Liu, *Sentiment analysis and opinion mining*, Springer Nature, 2022.
- [22] Z. Lin, X. Jin, X. Xu, W. Wang, X. Cheng, Y. Wang, A cross-lingual joint aspect/sentiment model for sentiment analysis, in: *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM '14*, Association for Computing Machinery, New York, NY, USA, 2014, p. 1089–1098. URL: <https://doi.org/10.1145/2661829.2662019>. doi:10.1145/2661829.2662019.
- [23] R. Klinger, P. Cimiano, Instance selection improves cross-lingual model training for fine-grained sentiment analysis, in: *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, Association for Computational Linguistics, Beijing, China, 2015, pp. 153–163. URL: <https://aclanthology.org/K15-1016>. doi:10.18653/v1/K15-1016.

- [24] M. S. Akhtar, P. Sawant, S. Sen, A. Ekbal, P. Bhattacharyya, Improving word embedding coverage in less-resourced languages through multi-linguality and cross-linguality: A case study with aspect-based sentiment analysis, *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 18 (2018).
- [25] W. Wang, S. J. Pan, Transition-based adversarial network for cross-lingual aspect extraction., in: *IJCAI*, 2018, pp. 4475–4481.
- [26] S. Jebbara, P. Cimiano, Zero-shot cross-lingual opinion target extraction, in: J. Burstein, C. Doran, T. Solorio (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)*, Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 2486–2495. URL: <https://aclanthology.org/N19-1257>. doi:10.18653/v1/N19-1257.
- [27] K. T.-K. Phan, D. Ngoc Hao, D. V. Thin, N. Luu-Thuy Nguyen, Exploring zero-shot cross-lingual aspect-based sentiment analysis using pre-trained multilingual language models, in: *2021 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, 2021, pp. 1–6. doi:10.1109/MAPR53640.2021.9585242.
- [28] H. Wu, Z. Wang, F. Qing, S. Li, Reinforced transformer with cross-lingual distillation for cross-lingual aspect sentiment classification, *Electronics* 10 (2021) 270.
- [29] D. N. H. Dang Van Thin, Hung Quoc Ngo, N. L.-T. Nguyen, Exploring zero-shot and joint training cross-lingual strategies for aspect-based sentiment analysis based on contextualized multilingual language models, *Journal of Information and Telecommunication* 7 (2023) 121–143.
- [30] K. Nam, Kpc-cf: Korean aspect-based sentiment analysis via pseudo-classifier with corpus filtering for low resource society, in: *First International Conference on Addressing Socioethical Effects of Artificial Intelligence*, 2024.
- [31] K. Sattar, Q. Umer, D. G. Vasbieva, S. Chung, Z. Latif, C. Lee, A multi-layer network for aspect-based cross-lingual sentiment classification, *IEEE Access* 9 (2021) 133961–133973.
- [32] J. Szolomicka, J. Kocon, Multiaspectemo: Multilingual and language-agnostic aspect-based sentiment analysis, in: *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2022, pp. 443–450. doi:10.1109/ICDMW58026.2022.00065.
- [33] C. Wu, B. Ma, Z. Zhang, N. Deng, Y. He, Y. Xue, Evaluating zero-shot multilingual aspect-based sentiment analysis with large language models, 2024. URL: <https://arxiv.org/abs/2412.12564>. arXiv:2412.12564.
- [34] R. Klinger, P. Cimiano, The USAGE review corpus for fine grained multi lingual opinion analysis, in: N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, European Language Resources Association (ELRA), Reykjavik, Iceland, 2014, pp. 2211–2218. URL: http://www.lrec-conf.org/proceedings/lrec2014/pdf/85_Paper.pdf.
- [35] R. Agerri, M. Cuadros, S. Gaines, G. Rigau, Opener: Open polarity enhanced named entity recognition, *Procesamiento del Lenguaje Natural* (2013) 215–218.
- [36] D. Hyun, J. Cho, H. Yu, Building large-scale English and Korean datasets for aspect-level sentiment analysis in automotive domain, in: D. Scott, N. Bel, C. Zong (Eds.), *Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online)*, 2020, pp. 961–966. URL: <https://aclanthology.org/2020.coling-main.83>. doi:10.18653/v1/2020.coling-main.83.
- [37] M. Luo, H. Fei, B. Li, S. Wu, Q. Liu, S. Poria, E. Cambria, M.-L. Lee, W. Hsu, Panosent: A panoptic sextuple extraction benchmark for multimodal conversational aspect-based sentiment analysis, in: *Proceedings of the 32nd ACM International Conference on Multimedia, MM '24*, Association for Computing Machinery, New York, NY, USA, 2024, p. 7667–7676. URL: <https://doi.org/10.1145/3664647.3680705>. doi:10.1145/3664647.3680705.
- [38] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, SemEval-2014 task 4: Aspect based sentiment analysis, in: P. Nakov, T. Zesch (Eds.), *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Association for Computational Linguistics, Dublin, Ireland, 2014, pp. 27–35. URL: <https://aclanthology.org/S14-2004>. doi:10.3115/v1/S14-2004.
- [39] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, I. Androutsopoulos, SemEval-2015 task 12: Aspect based sentiment analysis, in: P. Nakov, T. Zesch, D. Cer, D. Jurgens (Eds.), *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, Association for Computational Linguistics, Denver, Colorado, 2015, pp. 486–495. URL: <https://aclanthology.org/S15-2082>. doi:10.18653/v1/S15-2082.
- [40] Z. Fan, Z. Wu, X.-Y. Dai, S. Huang, J. Chen, Target-oriented opinion words extraction with target-fused neural sequence labeling, in: J. Burstein, C. Doran, T. Solorio (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)*, Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 2509–2518. URL: <https://aclanthology.org/N19-1259>. doi:10.18653/v1/N19-1259.
- [41] L. Xu, H. Li, W. Lu, L. Bing, Position-aware tagging for aspect sentiment triplet extraction, in: B. Webber, T. Cohn, Y. He, Y. Liu (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, Online, 2020, pp. 2339–2349. URL: <https://aclanthology.org/2020.emnlp-main.183>. doi:10.18653/v1/2020.emnlp-main.183.
- [42] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, L. Si, Knowing what, how and why: A near complete solution for aspect-based sentiment analysis, *Proceedings of the AAAI Conference on Artificial Intelligence* 34 (2020) 8600–8607.
- [43] H. Cai, R. Xia, J. Yu, Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions, in: C. Zong, F. Xia, W. Li, R. Navigli (Eds.), *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)*, Association for Computational Linguistics, Online, 2021, pp. 340–350. URL: <https://aclanthology.org/2021.acl-long.29>. doi:10.18653/v1/2021.acl-long.29.
- [44] W. Zhang, Y. Deng, X. Li, Y. Yuan, L. Bing, W. Lam, Aspect sentiment quad prediction as paraphrase generation, in: M.-F. Moens, X. Huang, L. Specia, S. W.-t. Yih (Eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Online and Punta Cana, Dominican Republic*, 2021, pp. 9209–9219. URL: <https://aclanthology.org/2021.emnlp-main.726>. doi:10.18653/v1/2021.emnlp-main.726.

- [45] J. Steinberger, T. Brychcín, M. Konkol, Aspect-level sentiment analysis in Czech, in: A. Balahur, E. van der Goot, R. Steinberger, A. Montoyo (Eds.), Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Association for Computational Linguistics, Baltimore, Maryland, 2014, pp. 24–30. URL: <https://aclanthology.org/W14-2605>. doi:10.3115/v1/W14-2605.
- [46] T. Hercig, T. Brychcín, L. Svoboda, M. Konkol, J. Steinberger, Unsupervised methods to improve aspect-based sentiment analysis in czech, *Computación y Sistemas* 20 (2016) 365–375.
- [47] J. Šmíd, P. Přibáň, O. Prazak, P. Kral, Czech dataset for complex aspect-based sentiment analysis tasks, in: N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, N. Xue (Eds.), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), ELRA and ICCL, Torino, Italia, 2024, pp. 4299–4310. URL: <https://aclanthology.org/2024.lrec-main.384>.
- [48] M. S. Akhtar, A. Ekbal, P. Bhattacharyya, Aspect based sentiment analysis in Hindi: Resource creation and evaluation, in: N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis (Eds.), Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), European Language Resources Association (ELRA), Portorož, Slovenia, 2016, pp. 2703–2709. URL: <https://aclanthology.org/L16-1429>.
- [49] Y. Nakayama, K. Murakami, G. Kumar, S. Bhingardive, I. Hardaway, A large-scale Japanese dataset for aspect-based sentiment analysis, in: N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, S. Piperidis (Eds.), Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 7014–7021. URL: <https://aclanthology.org/2022.lrec-1.758>.
- [50] J. Wang, C. Sun, S. Li, X. Liu, L. Si, M. Zhang, G. Zhou, Aspect sentiment classification towards question-answering with reinforced bidirectional attention network, in: A. Korhonen, D. Traum, L. Márquez (Eds.), Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 3548–3557. URL: <https://aclanthology.org/P19-1345>. doi:10.18653/v1/P19-1345.
- [51] J. Bu, L. Ren, S. Zheng, Y. Yang, J. Wang, F. Zhang, W. Wu, ASAP: A Chinese review dataset towards aspect category sentiment analysis and rating prediction, in: K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, Y. Zhou (Eds.), Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Online, 2021, pp. 2069–2079. URL: <https://aclanthology.org/2021.naacl-main.167>. doi:10.18653/v1/2021.naacl-main.167.
- [52] M. Saeidi, G. Bouchard, M. Liakata, S. Riedel, SentiHood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods, in: Y. Matsumoto, R. Prasad (Eds.), Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, The COLING 2016 Organizing Committee, Osaka, Japan, 2016, pp. 1546–1556. URL: <https://aclanthology.org/C16-1146>.
- [53] Q. Jiang, L. Chen, R. Xu, X. Ao, M. Yang, A challenge dataset and effective models for aspect-based sentiment analysis, in: K. Inui, J. Jiang, V. Ng, X. Wan (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 6280–6285. URL: <https://aclanthology.org/D19-1654>. doi:10.18653/v1/D19-1654.
- [54] X. Xing, Z. Jin, D. Jin, B. Wang, Q. Zhang, X. Huang, Tasty burgers, soggy fries: Probing aspect robustness in aspect-based sentiment analysis, in: B. Webber, T. Cohn, Y. He, Y. Liu (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 3594–3605. URL: <https://aclanthology.org/2020.emnlp-main.292>. doi:10.18653/v1/2020.emnlp-main.292.
- [55] K. Fukushima, Neocognitron, *Scholarpedia* 2 (2007) 1717.
- [56] S. Hochreiter, Long short-term memory, *Neural Computation* MIT-Press (1997).
- [57] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, I. Polosukhin, Attention is all you need, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, volume 30, Curran Associates, Inc., 2017, p. 6000–6010. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
- [58] J. D. Lafferty, A. McCallum, F. C. N. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, in: Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2001, p. 282–289.
- [59] E. F. Tjong Kim Sang, J. Veenstra, Representing text chunks, in: H. S. Thompson, A. Lascarides (Eds.), Ninth Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, Bergen, Norway, 1999, pp. 173–179. URL: <https://aclanthology.org/E99-1023>.
- [60] W. Zhang, X. Li, Y. Deng, L. Bing, W. Lam, Towards generative aspect-based sentiment analysis, in: C. Zong, F. Xia, W. Li, R. Navigli (Eds.), 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), Association for Computational Linguistics, Online, 2021, pp. 504–510. URL: <https://aclanthology.org/2021.acl-short.64>. doi:10.18653/v1/2021.acl-short.64.
- [61] Y. Mao, Y. Shen, J. Yang, X. Zhu, L. Cai, Seq2Path: Generating sentiment tuples as paths of a tree, in: S. Muresan, P. Nakov, A. Villavicencio (Eds.), Findings of the Association for Computational Linguistics: ACL 2022, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 2215–2225. URL: <https://aclanthology.org/2022.findings-acl.174>. doi:10.18653/v1/2022.findings-acl.174.
- [62] Z. Gou, Q. Guo, Y. Yang, MvP: Multi-view prompting improves aspect sentiment tuple prediction, in: A. Rogers, J. Boyd-Graber, N. Okazaki (Eds.), Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 4380–4397. URL: <https://aclanthology.org/2023.acl-long.240>. doi:10.18653/v1/2023.acl-long.240.
- [63] L. Xianlong, M. Yang, Y. Wang, Tagging-assisted generation model with encoder and decoder supervision for aspect sentiment triplet extraction, in: H. Bouamor, J. Pino, K. Bali (Eds.), Proceedings of the 2023 Conference on Empirical Methods in Natural Language

- Processing, Association for Computational Linguistics, Singapore, 2023, pp. 2078–2093. URL: <https://aclanthology.org/2023.emnlp-main.129>. doi:10.18653/v1/2023.emnlp-main.129.
- [64] J. Šmíd, P. Přibán, Prompt-based approach for Czech sentiment analysis, in: R. Mitkov, G. Angelova (Eds.), Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing, INCOMA Ltd., Shoumen, Bulgaria, Varna, Bulgaria, 2023, pp. 1110–1120. URL: <https://aclanthology.org/2023.ranlp-1.118>.
- [65] W. Zhang, Y. Deng, B. Liu, S. Pan, L. Bing, Sentiment analysis in the era of large language models: A reality check, in: K. Duh, H. Gomez, S. Bethard (Eds.), Findings of the Association for Computational Linguistics: NAACL 2024, Association for Computational Linguistics, Mexico City, Mexico, 2024, pp. 3881–3906. URL: <https://aclanthology.org/2024.findings-naacl.246>. doi:10.18653/v1/2024.findings-naacl.246.
- [66] J. Šmíd, P. Přibán, P. Kral, LLaMA-based models for aspect-based sentiment analysis, in: O. De Clercq, V. Barriere, J. Barnes, R. Klinger, J. Sedoc, S. Tafreshi (Eds.), Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 63–70. URL: <https://aclanthology.org/2024.wassa-1.6>.
- [67] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K. Weinberger (Eds.), Advances in Neural Information Processing Systems, volume 26, Curran Associates, Inc., 2013, p. 3111–3119. URL: https://proceedings.neurips.cc/paper_files/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf.
- [68] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, Enriching word vectors with subword information, Transactions of the Association for Computational Linguistics 5 (2017) 135–146.
- [69] J. Pennington, R. Socher, C. Manning, GloVe: Global vectors for word representation, in: A. Moschitti, B. Pang, W. Daelemans (Eds.), Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, 2014, pp. 1532–1543. URL: <https://aclanthology.org/D14-1162>. doi:10.3115/v1/D14-1162.
- [70] T. Luong, H. Pham, C. D. Manning, Bilingual word representations with monolingual quality in mind, in: P. Blunsom, S. Cohen, P. Dhillon, P. Liang (Eds.), Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, Association for Computational Linguistics, Denver, Colorado, 2015, pp. 151–159. URL: <https://aclanthology.org/W15-1521>. doi:10.3115/v1/W15-1521.
- [71] P. Koehn, Europarl: A parallel corpus for statistical machine translation, in: Proceedings of Machine Translation Summit X: Papers, Phuket, Thailand, 2005, pp. 79–86. URL: <https://aclanthology.org/2005.mtsummit-papers.11>.
- [72] S. Ruder, I. Vulić, A. Søgaard, A survey of cross-lingual word embedding models, J. Artif. Int. Res. 65 (2019) 569–630.
- [73] C. Dyer, V. Chahuneau, N. A. Smith, A simple, fast, and effective reparameterization of IBM model 2, in: L. Vanderwende, H. Daumé III, K. Kirchhoff (Eds.), Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Atlanta, Georgia, 2013, pp. 644–648. URL: <https://aclanthology.org/N13-1073>.
- [74] P. Koehn, An experimental management system, Prague Bulletin of Mathematical Linguistics 94 (2010) 87–96.
- [75] L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, C. Raffel, mT5: A massively multilingual pre-trained text-to-text transformer, in: K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, Y. Zhou (Eds.), Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Online, 2021, pp. 483–498. URL: <https://aclanthology.org/2021.naacl-main.41>. doi:10.18653/v1/2021.naacl-main.41.
- [76] S. Wu, M. Dredze, Beto, bentz, becas: The surprising cross-lingual effectiveness of BERT, in: K. Inui, J. Jiang, V. Ng, X. Wan (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 833–844. URL: <https://aclanthology.org/D19-1077>. doi:10.18653/v1/D19-1077.
- [77] Z. Wang, S. Mayhew, D. Roth, et al., Cross-lingual ability of multilingual bert: An empirical study, arXiv preprint arXiv:1912.07840 (2019).
- [78] T. Pires, E. Schlinger, D. Garrette, How multilingual is multilingual BERT?, in: A. Korhonen, D. Traum, L. Màrquez (Eds.), Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 4996–5001. URL: <https://aclanthology.org/P19-1493>. doi:10.18653/v1/P19-1493.
- [79] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, W. Wang, Language-agnostic BERT sentence embedding, in: S. Muresan, P. Nakov, A. Villavicencio (Eds.), Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 878–891. URL: <https://aclanthology.org/2022.acl-long.62/>. doi:10.18653/v1/2022.acl-long.62.
- [80] J. Pfeiffer, I. Vulić, I. Gurevych, S. Ruder, MAD-X: An Adapter-Based Framework for Multi-Task Cross-Lingual Transfer, in: B. Webber, T. Cohn, Y. He, Y. Liu (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 7654–7673. URL: <https://aclanthology.org/2020.emnlp-main.617>. doi:10.18653/v1/2020.emnlp-main.617.
- [81] Y. Ma, H. Peng, E. Cambria, Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm, Proceedings of the AAAI Conference on Artificial Intelligence 32 (2018).
- [82] B. Liang, H. Su, L. Gui, E. Cambria, R. Xu, Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks, Knowledge-Based Systems 235 (2022) 107643.
- [83] G. D’Aniello, M. Gaeta, I. La Rocca, Knowmis-absa: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis, Artificial Intelligence Review 55 (2022) 5543–5574.
- [84] T. Gao, J. Fang, H. Liu, Z. Liu, C. Liu, P. Liu, Y. Bao, W. Yan, LEGO-ABSA: A prompt-based task assemblable unified generative framework for multi-task aspect-based sentiment analysis, in: N. Calzolari, C.-R. Huang, H. Kim, J. Pustejovsky, L. Wanner, K.-S. Choi, P.-M. Ryu, H.-H. Chen, L. Donatelli, H. Ji, S. Kurohashi, P. Paggio, N. Xue, S. Kim, Y. Hahm, Z. He, T. K. Lee, E. Santus, F. Bond, S.-H. Na (Eds.),

- Proceedings of the 29th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 2022, pp. 7002–7012. URL: <https://aclanthology.org/2022.coling-1.610>.
- [85] M. Hu, Y. Wu, H. Gao, Y. Bai, S. Zhao, Improving aspect sentiment quad prediction via template-order data augmentation, in: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022, pp. 7889–7900. URL: <https://aclanthology.org/2022.emnlp-main.538>. doi:10.18653/v1/2022.emnlp-main.538.
- [86] A. Tamchyna, O. Fiala, K. Veselovská, Czech aspect-based sentiment analysis: A new dataset and preliminary results., in: ITAT, 2015, pp. 95–99.
- [87] L. Lenc, T. Hercig, Neural networks for sentiment analysis in czech, in: B. Brejová (Ed.), Proceedings of the 16th ITAT: Slovenskočeský NLP workshop (SloNLP 2016), volume 1649 of *CEUR Workshop Proceedings*, Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, CreateSpace Independent Publishing Platform, Bratislava, Slovakia, 2016, pp. 48–55.
- [88] P. Přibáň, O. Pražák, Improving aspect-based sentiment with end-to-end semantic role labeling model, in: R. Mitkov, G. Angelova (Eds.), Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing, INCOMA Ltd., Shoumen, Bulgaria, Varna, Bulgaria, 2023, pp. 888–897. URL: <https://aclanthology.org/2023.ranlp-1.96>.
- [89] A. García-Pablos, M. Cuadros, G. Rigau, W2vlda: Almost unsupervised system for aspect based sentiment analysis, *Expert Systems with Applications* 91 (2018) 127–137.
- [90] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation, *J. Mach. Learn. Res.* 3 (2003) 993–1022.
- [91] J. He, A. Wumaier, Z. Kadeer, W. Sun, X. Xin, L. Zheng, A local and global context focus multilingual learning model for aspect-based sentiment analysis, *IEEE Access* 10 (2022) 84135–84146.
- [92] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, in: NIPS 2014 Workshop on Deep Learning, December 2014, 2014.
- [93] M. R. R. Rana, T. Ali, A. Nawaz, Exploring multilingual reviews for aspect-based sentiment analysis using lexicon and bert, *Sigma Journal of Engineering and Natural Sciences* 42 (2024) 1469–1479.
- [94] A. Wawer, Few-shot methods for aspect-level sentiment analysis, *Information* 15 (2024).
- [95] J. Snell, K. Swersky, R. Zemel, Prototypical networks for few-shot learning, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17, Curran Associates Inc., Red Hook, NY, USA, 2017, p. 4080–4090.
- [96] Y. Yang, A. Katiyar, Simple and effective few-shot named entity recognition with structured nearest neighbor learning, in: B. Webber, T. Cohn, Y. He, Y. Liu (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 6365–6375. URL: <https://aclanthology.org/2020.emnlp-main.516/>. doi:10.18653/v1/2020.emnlp-main.516.
- [97] S. Lim, T. Yun, J. Kim, J. Choi, T. Kim, Analysis of multi-source language training in cross-lingual transfer, in: L.-W. Ku, A. Martins, V. Srikumar (Eds.), Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 712–725. URL: <https://aclanthology.org/2024.acl-long.42/>. doi:10.18653/v1/2024.acl-long.42.
- [98] X. Wan, Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis, in: M. Lapata, H. T. Ng (Eds.), Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Honolulu, Hawaii, 2008, pp. 553–561. URL: <https://aclanthology.org/D08-1058>.
- [99] X. Wan, Co-training for cross-lingual sentiment classification, in: K.-Y. Su, J. Su, J. Wiebe, H. Li (Eds.), Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Association for Computational Linguistics, Suntec, Singapore, 2009, pp. 235–243. URL: <https://aclanthology.org/P09-1027>.
- [100] A. Balahur, M. Turchi, Multilingual sentiment analysis using machine translation?, in: A. Balahur, A. Montoyo, P. M. Barco, E. Boldrini (Eds.), Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis, Association for Computational Linguistics, Jeju, Korea, 2012, pp. 52–60. URL: <https://aclanthology.org/W12-3709>.
- [101] A. Balahur, M. Turchi, Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis, *Computer Speech & Language* 28 (2014) 56–75.
- [102] A. Eriguchi, M. Johnson, O. Firat, H. Kazawa, W. Macherey, Zero-shot cross-lingual classification using multilingual neural machine translation, 2018. [arXiv:1809.04686](https://arxiv.org/abs/1809.04686).
- [103] H. Ghorbel, Experiments in cross-lingual sentiment analysis in discussion forums, in: K. Aberer, A. Flache, W. Jager, L. Liu, J. Tang, C. Guéret (Eds.), *Social Informatics*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 138–151.
- [104] H. Zhou, L. Chen, F. Shi, D. Huang, Learning bilingual sentiment word embeddings for cross-language sentiment classification, in: C. Zong, M. Strube (Eds.), Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Beijing, China, 2015, pp. 430–440. URL: <https://aclanthology.org/P15-1042>. doi:10.3115/v1/P15-1042.
- [105] P. Přibáň, J. Šmíd, A. Mištera, P. Král, Linear transformations for cross-lingual sentiment analysis, in: P. Sojka, A. Horák, I. Kopeček, K. Pala (Eds.), *Text, Speech, and Dialogue - 25th International Conference, TSD 2022, Brno, Czech Republic, September 6-9, 2022, Proceedings, volume 13502 of Lecture Notes in Computer Science*, Springer, 2022, pp. 125–137. URL: https://doi.org/10.1007/978-3-031-16270-1_11. doi:10.1007/978-3-031-16270-1_11.
- [106] P. Přibáň, J. Šmíd, J. Steinberger, A. Mištera, A comparative study of cross-lingual sentiment analysis, *Expert Systems with Applications* 247 (2024) 123247.
- [107] J. Barnes, R. Klinger, S. Schulte im Walde, Bilingual sentiment embeddings: Joint projection of sentiment across languages, in: I. Gurevych, Y. Miyao (Eds.), Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 2483–2493. URL: <https://aclanthology.org/P18-1231/>. doi:10.18653/v1/P18-1231.

- [108] P. Přibáň, J. Steinberger, Are the multilingual models better? improving Czech sentiment with transformers, in: R. Mitkov, G. Angelova (Eds.), Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021), INCOMA Ltd., Held Online, 2021, pp. 1138–1149. URL: <https://aclanthology.org/2021.ranlp-1.128>.
- [109] G. Thakkar, N. M. Preradovic, M. Tadic, Multi-task learning for cross-lingual sentiment analysis, arXiv preprint arXiv:2212.07160 (2022).
- [110] C. Wang, M. Banko, Practical transformer-based multilingual text classification, in: Y.-b. Kim, Y. Li, O. Rambow (Eds.), Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Papers, Association for Computational Linguistics, Online, 2021, pp. 121–129. URL: <https://aclanthology.org/2021.naacl-industry.16/>. doi:10.18653/v1/2021.naacl-industry.16.
- [111] V. Barriere, A. Balahur, Improving sentiment analysis over non-English tweets using multilingual transformers and automatic translation for data-augmentation, in: D. Scott, N. Bel, C. Zong (Eds.), Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online), 2020, pp. 266–271. URL: <https://aclanthology.org/2020.coling-main.23>. doi:10.18653/v1/2020.coling-main.23.
- [112] C. F. Ho, K. L. Chean, T. M. Lim, Leveraging machine translation to enhance sentiment analysis on multilingual text, in: Proceedings of the 2024 13th International Conference on Software and Computer Applications, ICSCA '24, Association for Computing Machinery, New York, NY, USA, 2024, p. 242–248. URL: <https://doi.org/10.1145/3651781.3651819>. doi:10.1145/3651781.3651819.
- [113] P. K. Roy, Deep ensemble network for sentiment analysis in bi-lingual low-resource languages, ACM Trans. Asian Low-Resour. Lang. Inf. Process. 23 (2024).
- [114] A. Ilyas, K. Shahzad, M. Kamran Malik, Emotion detection in code-mixed roman urdu - english text, ACM Trans. Asian Low-Resour. Lang. Inf. Process. 22 (2023).
- [115] M. Nazir, C. Faisal, M. A. Habib, H. Ahmad, Leveraging multilingual transformer for multiclass sentiment analysis in code-mixed data of low-resource languages, IEEE Access PP (2025).
- [116] G. Winata, S. Wu, M. Kulkarni, T. Solorio, D. Preotiuc-Pietro, Cross-lingual few-shot learning on unseen languages, in: Y. He, H. Ji, S. Li, Y. Liu, C.-H. Chang (Eds.), Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online only, 2022, pp. 777–791. URL: <https://aclanthology.org/2022.aac1-main.59/>. doi:10.18653/v1/2022.aac1-main.59.
- [117] S. Minaee, T. Mikolov, N. Nikzad, M. Chenaghlu, R. Socher, X. Amatriain, J. Gao, Large language models: A survey, 2024. URL: <https://arxiv.org/abs/2402.06196>. arXiv:2402.06196.
- [118] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, H. W. Chung, C. Sutton, S. Gehrmann, P. Schuh, K. Shi, S. Tsvyashchenko, J. Maynez, A. Rao, P. Barnes, Y. Tay, N. Shazeer, V. Prabhakaran, E. Reif, N. Du, B. Hutchinson, R. Pope, J. Bradbury, J. Austin, M. Isard, G. Gur-Ari, P. Yin, T. Duke, A. Levskaya, S. Ghemawat, S. Dev, H. Michalewski, X. Garcia, V. Misra, K. Robinson, L. Fedus, D. Zhou, D. Ippolito, D. Luan, H. Lim, B. Zoph, A. Spiridonov, R. Sepassi, D. Dohan, S. Agrawal, M. Omernick, A. M. Dai, T. S. Pillai, M. Pellat, A. Lewkowycz, E. Moreira, R. Child, O. Polozov, K. Lee, Z. Zhou, X. Wang, B. Saeta, M. Diaz, O. Firat, M. Catasta, J. Wei, K. Meier-Hellstern, D. Eck, J. Dean, S. Petrov, N. Fiedel, Palm: scaling language modeling with pathways, J. Mach. Learn. Res. 24 (2024).
- [119] OpenAI, J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, R. Avila, I. Babuschkin, S. Balaji, V. Balcom, P. Baltescu, H. Bao, M. Bavarian, J. Belgum, I. Bello, J. Berdine, G. Bernadett-Shapiro, C. Berner, L. Bogdonoff, O. Boiko, M. Boyd, A.-L. Brakman, G. Brockman, T. Brooks, M. Brundage, K. Button, T. Cai, R. Campbell, A. Cann, B. Carey, C. Carlson, R. Carmichael, B. Chan, C. Chang, F. Chantzis, D. Chen, S. Chen, R. Chen, J. Chen, M. Chen, B. Chess, C. Cho, C. Chu, H. W. Chung, D. Cummings, J. Currier, Y. Dai, C. Decareaux, T. Degry, N. Deutsch, D. Deville, A. Dhar, D. Dohan, S. Dowling, S. Dunning, A. Ecoffet, A. Eleti, T. Eloundou, D. Farhi, L. Fedus, N. Felix, S. P. Fishman, J. Forte, I. Fulford, L. Gao, E. Georges, C. Gibson, V. Goel, T. Gogineni, G. Goh, R. Gontijo-Lopes, J. Gordon, M. Grafstein, S. Gray, R. Greene, J. Gross, S. S. Gu, Y. Guo, C. Hallacy, J. Han, J. Harris, Y. He, M. Heaton, J. Heidecke, C. Hesse, A. Hickey, W. Hickey, P. Hoeschele, B. Houghton, K. Hsu, S. Hu, X. Hu, J. Huizinga, S. Jain, S. Jain, J. Jang, A. Jiang, R. Jiang, H. Jin, D. Jin, S. Jomoto, B. Jonn, H. Jun, T. Kaftan, Łukasz Kaiser, A. Kamali, I. Kanitscheider, N. S. Keskar, T. Khan, L. Kilpatrick, J. W. Kim, C. Kim, Y. Kim, J. H. Kirchner, J. Kiros, M. Knight, D. Kokotajlo, Łukasz Kondraciuk, A. Kondrich, A. Konstantinidis, K. Kosic, G. Krueger, V. Kuo, M. Lampe, I. Lan, T. Lee, J. Leike, J. Leung, D. Levy, C. M. Li, R. Lim, M. Lin, S. Lin, M. Litwin, T. Lopez, R. Lowe, P. Lue, A. Makanju, K. Malfacini, S. Manning, T. Markov, Y. Markovski, B. Martin, K. Mayer, A. Mayne, B. McGrew, S. M. McKinney, C. McLeavey, P. McMillan, J. McNeil, D. Medina, A. Mehta, J. Menick, L. Metz, A. Mishchenko, P. Mishkin, V. Monaco, E. Morikawa, D. Mossing, T. Mu, M. Murati, O. Murk, D. Mély, A. Nair, R. Nakano, R. Nayak, A. Neelakantan, R. Ngo, H. Noh, L. Ouyang, C. O’Keefe, J. Pachocki, A. Paino, J. Palermo, A. Pantuliano, G. Parascandolo, J. Parish, E. Parparita, A. Passos, M. Pavlov, A. Peng, A. Perelman, F. de Avila Belbute Peres, M. Petrov, H. P. de Oliveira Pinto, Michael, Pokorny, M. Pokrass, V. H. Pong, T. Powell, A. Power, B. Power, E. Proehl, R. Puri, A. Radford, J. Rae, A. Ramesh, C. Raymond, F. Real, K. Rimbach, C. Ross, B. Rotsted, H. Roussez, N. Ryder, M. Saltarelli, T. Sanders, S. Santurkar, G. Sastry, H. Schmidt, D. Schnurr, J. Schulman, D. Selsam, K. Sheppard, T. Sherbakov, J. Shieh, S. Shoker, P. Shyam, S. Sidor, E. Sigler, M. Simens, J. Sitkin, K. Slama, I. Sohl, B. Sokolowsky, Y. Song, N. Staudacher, F. P. Such, N. Summers, I. Sutskever, J. Tang, N. Tezak, M. B. Thompson, P. Tillet, A. Tootoonchian, E. Tseng, P. Tuggle, N. Turley, J. Tworek, J. F. C. Uribe, A. Vallone, A. Vijayvergiya, C. Voss, C. Wainwright, J. J. Wang, A. Wang, B. Wang, J. Ward, J. Wei, C. Weinmann, A. Welihinda, P. Welinder, J. Weng, L. Weng, M. Wiethoff, D. Willner, C. Winter, S. Wolrich, H. Wong, L. Workman, S. Wu, J. Wu, M. Wu, K. Xiao, T. Xu, S. Yoo, K. Yu, Q. Yuan, W. Zaremba, R. Zellers, C. Zhang, M. Zhang, S. Zhao, T. Zheng, J. Zhuang, W. Zhuk, B. Zoph, Gpt-4 technical report, 2024. URL: <https://arxiv.org/abs/2303.08774>. arXiv:2303.08774.
- [120] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, G. Lample, Llama: Open and efficient foundation language models, 2023. URL: <https://arxiv.org/abs/2302.13971>. arXiv:2302.13971.
- [121] H. Touvron, L. Martin, C. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, D. Bikel, L. Blecher, C. C. Ferrer, M. Chen, G. Cucurull, D. Esiobu, J. Fernandes, J. Fu, W. Fu, B. Fuller, C. Gao, V. Goswami, N. Goyal, A. Hartshorn, S. Hosseini,

- R. Hou, H. Inan, M. Kardaş, V. Kerkez, M. Khabisa, I. Kloumann, A. Korenev, P. S. Koura, M.-A. Lachaux, T. Lavril, J. Lee, D. Liskovich, Y. Lu, Y. Mao, X. Martinet, P. Mihaylov, P. Mishra, I. Molybog, Y. Nie, A. Poulton, J. Reizenstein, R. Rungta, K. Saladi, A. Schelten, R. Silva, E. M. Smith, R. Subramanian, X. E. Tan, B. Tang, R. Taylor, A. Williams, J. X. Kuan, P. Xu, Z. Yan, I. Zarov, Y. Zhang, A. Fan, M. Kambadur, S. Narang, A. Rodriguez, R. Stojnic, S. Edunov, T. Scialom, Llama 2: Open foundation and fine-tuned chat models, 2023. URL: <https://arxiv.org/abs/2307.09288>. arXiv:2307.09288.
- [122] A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, et al., The llama 3 herd of models, 2024. URL: <https://arxiv.org/abs/2407.21783>. arXiv:2407.21783.
- [123] Q. Zhong, L. Ding, J. Liu, B. Du, D. Tao, Can chatgpt understand too? a comparative study on chatgpt and fine-tuned bert, 2023. URL: <https://arxiv.org/abs/2302.10198>. arXiv:2302.10198.
- [124] P. F. Simmering, P. Huoviala, Large language models for aspect-based sentiment analysis, 2023. URL: <https://arxiv.org/abs/2310.18025>. arXiv:2310.18025.
- [125] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Q. V. Le, D. Zhou, Chain-of-thought prompting elicits reasoning in large language models, in: Proceedings of the 36th International Conference on Neural Information Processing Systems, NIPS '22, Curran Associates Inc., Red Hook, NY, USA, 2024.
- [126] H. Fei, B. Li, Q. Liu, L. Bing, F. Li, T.-S. Chua, Reasoning implicit sentiment with chain-of-thought prompting, in: A. Rogers, J. Boyd-Graber, N. Okazaki (Eds.), Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 1171–1182. URL: <https://aclanthology.org/2023.acl-short.101/>. doi:10.18653/v1/2023.acl-short.101.
- [127] T. Filip, M. Pavlíček, P. Sosík, Fine-tuning multilingual language models in twitter/x sentiment analysis: a study on eastern-european v4 languages, 2024. URL: <https://arxiv.org/abs/2408.02044>. arXiv:2408.02044.
- [128] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. L. Scao, T. Lavril, T. Wang, T. Lacroix, W. E. Sayed, Mistral 7b, 2023. URL: <https://arxiv.org/abs/2310.06825>. arXiv:2310.06825.
- [129] N. Mughal, G. Mujtaba, S. Shaikh, A. Kumar, S. M. Daudpota, Comparative analysis of deep natural networks and large language models for aspect-based sentiment analysis, IEEE Access 12 (2024) 60943–60959.
- [130] P. He, X. Liu, J. Gao, W. Chen, Deberta: Decoding-enhanced bert with disentangled attention, 2021. URL: <https://arxiv.org/abs/2006.03654>. arXiv:2006.03654.
- [131] B. Ding, C. Qin, R. Zhao, T. Luo, X. Li, G. Chen, W. Xia, J. Hu, A. T. Luu, S. Joty, Data augmentation using LLMs: Data perspectives, learning paradigms and challenges, in: L.-W. Ku, A. Martins, V. Srikumar (Eds.), Findings of the Association for Computational Linguistics: ACL 2024, Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 1679–1705. URL: <https://aclanthology.org/2024.findings-acl.97>. doi:10.18653/v1/2024.findings-acl.97.
- [132] Z. Li, W. Chen, S. Li, H. Wang, J. Qian, X. Yan, Controllable dialogue simulation with in-context learning, in: Y. Goldberg, Z. Kozareva, Y. Zhang (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2022, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022, pp. 4330–4347. URL: <https://aclanthology.org/2022.findings-emnlp.318>. doi:10.18653/v1/2022.findings-emnlp.318.
- [133] A. G. Møller, A. Pera, J. Dalsgaard, L. Aiello, The parrot dilemma: Human-labeled vs. LLM-augmented data in classification tasks, in: Y. Graham, M. Purver (Eds.), Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 2: Short Papers), Association for Computational Linguistics, St. Julian's, Malta, 2024, pp. 179–192. URL: <https://aclanthology.org/2024.eacl-short.17>.
- [134] Q. Zhong, H. Li, L. Zhuang, J. Liu, B. Du, Iterative data generation with large language models for aspect-based sentiment analysis, 2024. URL: <https://arxiv.org/abs/2407.00341>. arXiv:2407.00341.
- [135] T. Detmers, A. Pagnoni, A. Holtzman, L. Zettlemoyer, Qlora: Efficient finetuning of quantized llms, in: A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, S. Levine (Eds.), Advances in Neural Information Processing Systems, volume 36, Curran Associates, Inc., 2023, pp. 10088–10115. URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/1feb87871436031bdc0f2beaa62a049b-Paper-Conference.pdf.
- [136] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, LoRA: Low-rank adaptation of large language models, in: International Conference on Learning Representations, 2022. URL: <https://openreview.net/forum?id=nZeVKeeFYf9>.
- [137] C. Liu, W. Zhang, Y. Zhao, A. T. Luu, L. Bing, Is translation all you need? a study on solving multilingual tasks with large language models, 2024. URL: <https://arxiv.org/abs/2403.10258>. arXiv:2403.10258.
- [138] A. Üstün, V. Aryabumi, Z. Yong, W.-Y. Ko, D. D'souza, G. Onilude, N. Bhandari, S. Singh, H.-L. Ooi, A. Kayid, F. Vargus, P. Blunsom, S. Longpre, N. Muennighoff, M. Fadaee, J. Kreutzer, S. Hooker, Aya model: An instruction finetuned open-access multilingual language model, in: L.-W. Ku, A. Martins, V. Srikumar (Eds.), Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 15894–15939. URL: <https://aclanthology.org/2024.acl-long.845/>. doi:10.18653/v1/2024.acl-long.845.
- [139] V. Aryabumi, J. Dang, D. Talupuru, S. Dash, D. Cairuz, H. Lin, B. Venkitesh, M. Smith, J. A. Campos, Y. C. Tan, K. Marchisio, M. Bartolo, S. Ruder, A. Locatelli, J. Kreutzer, N. Frosst, A. Gomez, P. Blunsom, M. Fadaee, A. Üstün, S. Hooker, Aya 23: Open weight releases to further multilingual progress, 2024. URL: <https://arxiv.org/abs/2405.15032>. arXiv:2405.15032.
- [140] A. Mitra, L. D. Corro, S. Mahajan, A. Codaş, C. Simoes, S. Agarwal, X. Chen, A. Razdaibiedina, E. Jones, K. Aggarwal, H. Palangi, G. Zheng, C. Rosset, H. Khanpour, A. Awadallah, Orca 2: Teaching small language models how to reason, 2023. URL: <https://arxiv.org/abs/2311.11045>. arXiv:2311.11045.
- [141] J. Šmíd, Cross-lingual Aspect-Based Sentiment Analysis, Master's thesis, University of West Bohemia, Faculty of Applied Sciences, Plzeň, 2023.

- [142] B. Ding, C. Qin, R. Zhao, T. Luo, X. Li, G. Chen, W. Xia, J. Hu, A. T. Luu, S. Joty, Data augmentation using large language models: Data perspectives, learning paradigms and challenges, 2024. URL: <https://arxiv.org/abs/2403.02990>. arXiv:2403.02990.
- [143] J. Zhang, R. Xie, Y. Hou, X. Zhao, L. Lin, J.-R. Wen, Recommendation as instruction following: A large language model empowered recommendation approach, ACM Trans. Inf. Syst. (2024). Just Accepted.