A Survey on Cross-Lingual Summarization

Jiaan Wang[1*], **Fandong Meng**[2†], **Duo Zheng**[4], **Yunlong Liang**[2]
Zhixu Li[3†], **Jianfeng Qu**[1] **and Jie Zhou**[2]

[1]School of Computer Science and Technology, Soochow University, Suzhou, China
[2]Pattern Recognition Center, WeChat AI, Tencent Inc, China
[3]Shanghai Key Laboratory of Data Science, School of Computer Science,
Fudan University, Shanghai, China
[4]Beijing University of Posts and Telecommunications, Beijing, China
jawang1@stu.suda.edu.cn, {fandongmeng,yunlonliang,withtomzhou}@tencent.com
zd@bupt.edu.cn, zhixuli@fudan.edu.cn, jfqu@suda.edu.cn

Abstract

Cross-lingual summarization is the task of generating a summary in one language (e.g., English) for the given document(s) in a different language (e.g., Chinese). Under the globalization background, this task has attracted increasing attention of the computational linguistics community. Nevertheless, there still remains a lack of comprehensive review for this task. Therefore, we present the first systematic critical review on the datasets, approaches, and challenges in this field. Specifically, we carefully organize existing datasets and approaches according to different construction methods and solution paradigms, respectively. For each type of dataset or approach, we thoroughly introduce and summarize previous efforts and further compare them with each other to provide deeper analyses. In the end, we also discuss promising directions and offer our thoughts to facilitate future research. This survey is for both beginners and experts in cross-lingual summarization, and we hope it will serve as a starting point as well as a source of new ideas for researchers and engineers interested in this area.

1 Introduction

To help people efficiently grasp the gist of documents in a foreign language, Cross-Lingual Summarization (XLS) aims to generate a summary in the target language from the given document(s) in a different source language. This task could be regarded as a combination of monolingual summarization (MS) and machine translation (MT), both of which are unsolved natural language processing (NLP) tasks and have been continuously studied for decades (Paice, 1990; Brown et al., 1993). XLS is an extremely challenging task: (1) from the perspective of data, unlike MS, naturally occurring documents in a source language paired with the corresponding summaries in different target languages are rare, making it difficult to collect large-scale and human-annotated datasets (Ladhak et al., 2020; Perez-Beltrachini and Lapata, 2021); (2) from the perspective of models, XLS requires both the abilities to translate and summarize, which makes it hard to generate accurate summaries by directly conducting XLS (Cao et al., 2020).

Despite its importance, XLS has attracted a little attention (Leuski et al., 2003; Wan et al., 2010) in the statistical learning era due to its difficulties and the scarcity of parallel corpus. Recent years have witnessed the rapid development of neural networks, especially the emergence of pre-trained encoder-decoder models (Zhang et al., 2020a; Raffel et al., 2020; Lewis et al., 2020; Liu et al., 2020; Tang et al., 2021; Xue et al., 2021), making neural summarizers and translators achieve impressive performance. Meanwhile, creating large-scale XLS datasets has proven feasible by utilizing existing MS datasets (Zhu et al., 2019; Wang et al., 2022b) or Internet resources (Ladhak et al., 2020; Perez-Beltrachini and Lapata, 2021). The aforementioned successes have laid the foundation for the XLS research field and gradually attracted interest in XLS. In particular, recent researchers put their efforts into solving the XLS task and published more than 20 papers over the past five years. Nevertheless,

[*]Work was done when Jiaan Wang was interning at Pattern Recognition Center, WeChat AI, Tencent Inc, China.
[†]Corresponding authors.

Transactions of the Association for Computational Linguistics, vol. 10, pp. 1304–1323, 2022. https://doi.org/10.1162/tacl_a_00520
Action Editor: Yang Liu. Submission batch: 4/2022; Revision batch: 7/2022; Published 11/2022.

there still lacks a systematic review of progresses, challenges, and opportunities of XLS.

To fill the above gap and help new researchers, in this paper we provide the first comprehensive review of existing efforts relevant to XLS and give multiple promising directions for future research. Specifically, we first briefly introduce the formal definition and evaluation metrics of XLS (§ 2), which serves as a strong background before delving further into XLS. Then, we provide an exhaustive overview of existing XLS research datasets (§ 3). In detail, to alleviate the scarcity of XLS data, previous work resorts to different ways to construct large-scale benchmark datasets, which are divided into synthetic datasets and multi-lingual website datasets. The synthetic datasets (Zhu et al., 2019; Bai et al., 2021a; Wang et al., 2022b) are constructed through (manually or automatically) translating the summaries of existing MS datasets from a source language to target languages while the multi-lingual website datasets (Nguyen and Daumé III, 2019; Ladhak et al., 2020; Fatima and Strube, 2021; Perez-Beltrachini and Lapata, 2021) are collected from websites that provide multi-lingual versions for their content.

Next, we thoroughly introduce and summarize existing models, which are organized with respect to different paradigms, namely, pipeline (§ 4) and end-to-end (§ 5). In detail, the pipeline models adopt either translate-then-summarize approaches (Leuski et al., 2003; Boudin et al., 2011; Wan, 2011; Yao et al., 2015; Zhang et al., 2016; Linhares Pontes et al., 2018; Wan et al., 2018; Ouyang et al., 2019) or summarize-then-translate approaches (Orăsan and Chiorean, 2008; Wan et al., 2010). In this manner, the pipeline models avoid conducting XLS directly, thus bypassing the model challenge we discussed previously. However, the pipeline method suffers from error propagation and recurring latency, making it not suitable for the real-world scenario (Ladhak et al., 2020). Consequently, the end-to-end method has attracted more attention. To alleviate the model challenge, it generally utilizes the related tasks (e.g., MS and MT) as auxiliaries or resorts to external resources. The end-to-end models mainly fall into four categories: multi-task methods (Zhu et al., 2019; Takase and Okazaki, 2020; Cao et al., 2020; Bai et al., 2021a; Liang et al., 2022), knowledge-distillation methods (Ayana et al., 2018; Duan et al., 2019; Nguyen and Luu, 2022), resource-enhanced methods (Zhu et al., 2020; Jiang et al., 2022), and pre-training methods (Dou et al., 2020; Xu et al., 2020; Ma et al., 2021; Chi et al., 2021a; Wang et al., 2022b). For each category, we will thoroughly go through the previous work and discuss the corresponding pros and cons. Finally, we also point out multiple promising directions on XLS to push forward the future research (§ 6), followed by conclusions (§ 7). Our contributions are concluded as follows:

- To the best of our knowledge, this survey is the first that presents a thorough review of XLS.
- We comprehensively review the existing XLS work and carefully organize them according to different frameworks.
- We suggest multiple promising directions to facilitate future research on XLS.

2 Background

2.1 Task Definition

Given a collection of documents in the source language $\mathcal{D} = \{D_i\}_{i=1}^{m}$ (m denotes the number of documents and $m \geq 1$), the goal of XLS is to generate the corresponding summary in the target language $Y = \{y_i\}_{i=1}^{n}$ with n words. The conditional distribution of XLS models is:

$$p_\theta(Y|\mathcal{D}) = \prod_{t=1}^{n} p_\theta(y_t|\mathcal{D}, y_{1:t-1})$$

where θ represents model parameters and $y_{1:t-1}$ is the partial ground truth summary.

It is worth noting that: (1) when $m > 1$, this task is upgraded to cross-lingual multi-document summarization (XLMS) which has been discussed by some previous studies (Orăsan and Chiorean, 2008; Boudin et al., 2011; Zhang et al., 2016); (2) when the given documents are dialogues, the task becomes cross-lingual dialogue summarization (XLDS) which has been recently proposed by Wang et al. (2022b). The XLMS and XLDS are also within the scope of this survey. Furthermore, we define the source and the target languages in XLS should be two *exactly distinct human languages*, which also means (1) if the source language is in code-mixed style of two natural

languages (e.g., Chinese and English), the target language should not be either of the both; (2) the programming languages (e.g., PYTHON or JAVA) should not be the source or the target language.[1]

2.2 Evaluation

Following MS, ROUGE scores (Lin, 2004) are universally adopted as the basic automatic metrics for XLS, especially the F1 scores of ROUGE-1, ROUGE-2, and ROUGE-L, which measure the unigram, bigram, and longest common sequence between the ground truth and the generated summaries, respectively. Nevertheless, the original ROUGE scores are specifically designed for English. To make these metrics suitable for other languages, some useful toolkits have been released, for example, `multi-lingual ROUGE`[2] and `MLROUGE`.[3] In addition to these metrics based on lexical overlap, recent work proposes new metrics based on the semantic similarity (token/word embeddings), such as `MoverScore`[4] (Zhao et al., 2019) and `BERTScore`[5] (Zhang et al., 2020b), whose great consistency with human judgements on MS has been shown (Koto et al., 2021).

3 Datasets

In this section, we review available large-scale XLS datasets[6] and further divide them into two categories: synthetic datasets (§ 3.1) and multilingual website datasets (§ 3.2). For each category, we will introduce the construction details and the key characteristics of the corresponding datasets. In addition, we compare these two categories to provide a deeper understanding (§ 3.3).

[1] If the source language is a programming language while the target language is a human language, the task becomes code summarization, which is beyond the scope of this survey.

[2] `https://github.com/csebuetnlp/xl-sum/tree/master/multilingual_rouge_scoring`.

[3] `https://github.com/dqwang122/MLROUGE`.

[4] `https://github.com/AIPHES/emnlp19-moverscore`.

[5] `https://github.com/Tiiiger/bert_score`.

[6] There are also some XLS datasets in the statistical learning era, e.g., multiple MultiLing datasets (Giannakopoulos, 2013; Giannakopoulos et al., 2015) and the translated DUC2001 dataset (Wan, 2011). However, these datasets are either not public or extremely limited in scale (typically less than 100 samples). Thus, we do not go into these datasets in depth.

3.1 Synthetic Datasets

Intuitively, one straightforward way to build XLS datasets is directly translating the summaries of a MS dataset from their original language to different target languages. The datasets built in this way are named synthetic datasets, which could benefit from existing MS datasets.

Dataset Construction. En2ZhSum (Zhu et al., 2019) is constructed through utilizing a sophisticated MT service[7] to translate the summaries of CNN/Dailymail (Hermann et al., 2015) and MSMO (Zhu et al., 2018) from English to Chinese. In the same way, Zh2EnSum (Zhu et al., 2019) is built through translating the summaries of LCSTS (Hu et al., 2015) from Chinese to English. Later, Bai et al. (2021a) propose En2DeSum through translating the English Gigaword[8] to German using the WMT'19 English-German winner MT model (Ng et al., 2019).

More recently, Wang et al. (2022b) construct XSAMSum and XMediaSum, which directly employ professional translators to translate summaries of two dialogue-oriented MS datasets, that is, SAMSum (Gliwa et al., 2019) and MediaSum (Zhu et al., 2021), from English to both German and Chinese. In this way, their datasets achieve much higher quality than those automatically constructed ones.

Quality Controlling. Since the translation results provided by MT services might contain flaws, En2ZhSum, Zh2EnSum, and En2DeSum further use the round-trip translation (RTT) strategy to filter out low-quality samples. Specifically, given a monolingual document-summary pair $\langle D_{\text{src}}, S_{\text{src}} \rangle$, the summary S_{src} is first translated to a target langauge S'_{tgt}, and then S'_{tgt} is translated back to the source language S'_{src}. Next, $\langle D_{src}, S'_{tgt} \rangle$ will be retained as an XLS sample only if the ROUGE scores between S_{src} and S'_{src} exceed the pre-defined thresholds. In addition, the translated summaries in the test set of En2ZhSum and Zh2EnSum are post-edited by human annotators to ensure the reliability of model evaluation.

As for manually translated synthetic datasets, namely, XSAMSum and XMediaSum, Wang et al. (2022b) design a quality control loop, where data

[7] `http://www.zkfy.com/`.

[8] LDC2011T07.

Dataset	Trans.	Genre	Scale	Src Lang.	Tgt Lang.
En2ZhSum	Auto.	News	371k	En	Zh
Zh2EnSum	Auto.	News	1.7M	Zh	En
En2DeSum	Auto.	News	438k	En	De
XSAMSum	Manu.	Dial.	16k×2	En	De/Zh
XMediaSum	Manu.	Dial.	40k×2	En	De/Zh

Table 1: Overview of existing synthetic XLS datasets. ''*Trans.*'' indicates the translation method (automatic or manual) to construct datasets. The ''*genres*'' of these datasets are divided into news articles and dialogues according to the basic MS datasets. For ''*scale*'', some datasets contain two cross-lingual directions, thus we use $\times 2$ to calculate the overall scale. ''*Src Lang.*'' and ''*Tgt Lang.*'' denote the source and target languages for each dataset, respectively (En: English, Zh: Chinese, and De: German).

reviewers and experts participate to ensure the accuracy of the translation.

Dataset Statistics. Table 1 compares previous synthetic datasets in terms of the translation method, genre, scale, source language, and target language. We conclude that: (1) There is a trade-off between scale and quality. In line with MS, the scale of XLS datasets in the news domain is much larger than others since news articles are convenient to collect. When faced with such large-scale datasets, it is expensive and even impractical to manually translate or post-edit all their summaries. Thus, these datasets generally adopt automatic translation methods, causing limited quality. (2) The XLS datasets in the dialogue domain are more challenging than those in the news domain. Besides the limited scale, the key information of one dialogue is often scattered and spans multiple utterances, leading to low information density (Feng et al., 2022c), which together with complex dialogue phenomena (e.g., coreference, repetition, and interruption) makes the task quite challenging (Wang et al., 2022b).

3.2 Multi-Lingual Website Datasets

In the globalization process, online resources across different languages are overwhelmingly growing. One reason is that many websites start to provide multi-lingual versions for their content to facilitate global users. Therefore, these websites might contain a large number of parallel documents in different languages. Some researchers try to utilize such resources to establish XLS datasets.

Dataset Construction. Nguyen and Daumé III (2019) collect news articles from the Global Voices website,[9] which reports and translates news about unheard voices across the globe. The translated news on this website is performed by volunteer translators. Each news article also links to its parallel articles in other languages, if available. Thus, it is convenient to obtain different language versions of an article. Then, they employ crowdworkers to write English summaries for hundreds of selected English articles. In this manner, the non-English articles together with the English summaries constitute the Global Voices XLS dataset.[10] Although this dataset utilizes online resources, the way to collect summaries (i.e., crowd-sourcing) limits its scale and directions (the target language must be English).

To alleviate the dilemma, WikiLingua (Ladhak et al., 2020) collects multi-lingual guides from WikiHow,[11] where each step in a guide consists of a paragraph and the corresponding one-sentence summary. Heuristically, the dataset combines paragraphs and one-sentence summaries of all the steps in one guide to create a monolingual article-summary pair. With the help of hyperlinks between parallel guides in different languages, the article in one language and its summary in another one are easy to align. In this way, WikiLingua collects articles and the corresponding summaries in 18 different languages, leading to 306 (18×17) directions. Similarly, Perez-Beltrachini and Lapata (2021) construct XLS datasets from Wikipedia,[12] a widely used multilingual encyclopedia. In detail, the Wikipedia articles are typically organized into lead sections and bodies. They focus on 4 languages and pair lead sections with the corresponding bodies in different languages to construct XLS samples. In the end, the collected samples form the XWikis dataset with 12 directions.

[9] https://globalvoices.org/.

[10] The Global Voices dataset contains gv-snippet and gv-crowd two subsets. The former cannot well meet the need of XLS due to its low quality (Nguyen and Daumé III, 2019), thus we only introduce the gv-crowd subset.

[11] https://www.wikihow.com/.

[12] https://www.wikipedia.org/.

Dataset	Domain	L	D	Scale (avg / max / min)
Global Voices	News	15	14	208 / 487 / 75
CrossSum	News	45	1936	845 / 45k / 1
WikiLingua	Guides	18	306	18k / 113k / 915
XWikis	Encyclopedia	4	12	214k / 469k / 52k

Table 2: Overview of representative multi-lingual website datasets. ''*L*'' denotes the number of languages involved in each dataset. ''*D*'' indicates the number of cross-lingual directions. ''*Scale (avg/max/min)*'' calculates the average/maximum/minimum number of XLS samples per direction.

Additionally, Hasan et al. (2021a) construct the CrossSum dataset by automatically aligning identical news articles written in different languages from the XL-Sum dataset (Hasan et al., 2021b). The multi-lingual news article-summary pairs in XL-Sum are collected from the BBC website.[13] As a result, CrossSum involves 45 languages and 1936 directions.

Quality Control. For the manually annotated dataset (i.e., Global Voices), Nguyen and Daumé III (2019) employ human evaluation to remove low-quality annotated summaries to ensure the quality. For automatically collected datasets (i.e., WikiLingua and XWikis), they typically extract the desired content from the websites via heuristic matching rules to ensure the correctness. As for the automatically aligned dataset (i.e., CrossSum), Hasan et al. (2021a) adopt LaBSE (Feng et al., 2022a) to encode all summaries from XL-Sum (Hasan et al., 2021b). Then, they align documents belonging to different languages based on the cosine similarity of corresponding summaries, and pre-define a minimum similarity score to reduce the number of incorrect alignments.

Dataset Statistics. Table 2 lists the key characteristics of the representative multi-lingual website datasets. It is worth noting that the number of XLS samples in each direction of the same dataset may be different since different articles might be available in different languages. Hence, we measure the overall scale of each dataset from its average, maximum, and minimum number of XLS samples per direction, respectively. We find that: (1) The scale of Global Voices is extremely lower than other datasets due to the different methods for collecting summaries. Specifically, the WikiLingua, XWikis and XL-Sum (the basis of CrossSum) datasets automatically extract a huge number of summaries from online resources via simple strategies rather than crowd-sourcing. (2) CrossSum and WikiLingua involve more languages than the others, and most language pairs have intersectional articles, resulting in numerous cross-lingual directions.

[13] https://www.bbc.com/.

3.3 Discussion

According to the above review of large-scale XLS datasets, the approaches for building datasets are summarized as: (I) manually or (II) automatically translating the summaries of MS datasets; (III) automatically collecting documents as well as summaries from multi-lingual websites.

Among them, approach I involves less noise than others since its translation and quality control are performed by professional translators rather than machine translation or volunteers. However, this approach is too labor-intensive and costly to build large-scale datasets. For instance, to control costs, XMediaSum (Wang et al., 2022b) only manually translates part of ($\sim$8.6%) summaries of MediaSum (Zhu et al., 2021). Besides, Zh2EnSum and En2ZhSum (Zhu et al., 2019) are automatically collected via approach II, and only their test sets have been manually corrected. Therefore, despite the high quality of the constructed data, approach I is more suitable for building validation and test sets of large-scale XLS datasets rather than the whole datasets.

Approaches II and III could be adopted to build whole XLS datasets. We discuss them in the following situations:

(1) High-resource source languages $\Rightarrow$ high-resource target languages: This situation has been well studied in previous work, and most of the proposed XLS datasets focus on this situation. Both approaches II and III are useful to construct XLS datasets whose source and target languages are both high-resource languages.

(2) High-resource source languages $\Rightarrow$ low-resource target languages: When the documents and summaries from XLS datasets are, respectively, in a high-resource language and a low-resource language, approach III loses its effectiveness. This is because, for a multi-lingual website, its content in a low-resource language is typically less than that in a high-resource

language. As a result, the number of collected XLS samples involving low-resource languages is significantly limited. For example, WikiLingua (Ladhak et al., 2020), as a multi-lingual website dataset, contains 113.2k English⇒Spanish samples, but only 7.2k English⇒Czech samples. In this situation, approach II might be a possible way to collect a large number of samples. Note that the MT from a high-resource language to a low-resource language might involve more translation flaws than those between two high-resource languages. Thus, besides the RTT strategy, how to filter out the potential flaws is worthy for further study.

(3) Low-resource source languages ⇒ high- or low-resource target languages: If the source language is low-resource, there might be no MS dataset and enough website content in this language, leading to the failures of approaches II and III. Therefore, how to build datasets in this situation is still an open-ended problem, which needs to be explored in the future. As pointed by Feng et al. (2022b), one straightforward approach is to automatically translate both documents and summaries from high-resource MS datasets. However, translating documents with hundreds of words might introduce substantial noise, especially when low-resource languages are involved. Thus, its practicality and reliability need more careful justification.

4 Pipeline Methods

Early XLS work generally focuses on the pipeline methods whose main idea is decomposing XLS into MS and MT sub-tasks, and then accomplishing them step by step. These methods can be further divided into summarize-then-translate (Sum-Trans) and translate-then-summarize (Trans-Sum) types according to the finished order of sub-tasks. For each type, we will systematically present previous methods. Additionally, we compare these two types to provide deeper analyses.

4.1 Sum-Trans

Orăsan and Chiorean (2008) utilize the Maximum Marginal Relevance (MMR) algorithm to summarize Romanian news, and then translate the summaries from Romanian to English via the eTranslator MT service.[14] Furthermore, Wan et al. (2010) find the translated summaries might fall into low readability due to the limited MT performance at that time. To alleviate this issue, they first use a trained SVM model (Cortes and Vapnik, 1995) to predict the translation quality of each English sentence, where the model only leverages features in the English sentences. Then, they select sentences with high quality and informativeness to form summaries which are finally translated to Chinese by Google MT service.[15]

4.2 Trans-Sum

Compared with Sum-Trans, Trans-Sum attracts more research attention, and this type of pipeline method can be further classified into three sub-types depending on whether its summarizer is extractive, compressive, or abstractive:

- The extractive method selects complete sentences from the translated documents as summaries.
- The compressive method first extracts key sentences from the translated documents, and further removes non-relevant or redundant words in the key sentences to obtain the final summaries.
- The abstractive method generates new sentences as summaries, which are not limited to original words or phrases.

Note that we do not classify the Sum-Trans approaches in the same manner since their summarizers are all extractive.

Extractive Trans-Sum. Leuski et al. (2003) build a cross-lingual information delivery system that first translates Hindi documents to English via a statistical MT model and then selects important English sentences to form summaries. In this system, the summarizer only uses the document information from the target language side, which heavily depends on the MT results and might lead to flawed summaries. However, semantic information from both sides should be taken into account.

To this end, after translating English documents to Chinese, Wan (2011) designs two graph-based summarizers (i.e., SimFusion and CoRank) which utilize bilingual information to output the final Chinese summaries: (i) the SimFusion

[14] https://www.etranslator.ro/.

[15] https://cloud.google.com/translate.

summarizer first measures the saliency scores of Chinese sentences through combing the English-side and Chinese-side similarity, and then, the salient Chinese sentences constitute the final summaries; (ii) the CoRank summarizer simultaneously ranks both English and Chinese sentences by incorporating mutual influences between them, and then, the top-ranking Chinese sentences are used to constitute summaries.

Later, Boudin et al. (2011) translate documents from English to French, and then use SVM regression method to predict translation quality of each sentence based on bilingual features. Next, the crucial translated sentences are selected based on a modified PageRank algorithm (Page et al., 1999) considering the translation quality. Lastly, the redundant sentences are removed from the selected sentences to form the final summaries.

Compressive Trans-Sum. Inspired by phrase-based MT, Yao et al. (2015) propose a compressive summarization method that simultaneously selects and compresses sentences. Specifically, the sentence selection is based on bilingual features, and the sentence compression is performed by removing the redundant or poorly translated phrases in one single sentence. To further excavate the complementary information of similar sentences, Zhang et al. (2016) first parse bilingual documents into predicate-argument structures (PAS), and then produce summaries by fusing bilingual PAS structures. In this way, several salient PAS elements (concepts or facts) from different sentences can be merged into one summary sentence. Similarly, Linhares Pontes et al. (2018) take bilingual lexical chunks into account during measuring the sentence similarity and further compress sentences at both single- and multi-sentence levels.

Abstractive Trans-Sum. With the emergence of large-scale synthetic XLS datasets (Zhu et al., 2019), researchers attempt to adopt the sequence-to-sequence models as summarizers in Trans-Sum methods. Considering that the translated documents might contain flaws, Ouyang et al. (2019) train an abstractive summarizer (i.e., PGNet, See et al. 2017) on English pairs of a noisy document and a clean summary. In this manner, the summarizer could achieve good robustness, when summarizing the English documents which are translated from a low-resource language.

4.3 Sum-Trans vs. Trans-Sum

We compare Sum-Trans and Trans-Sum in the following situations:

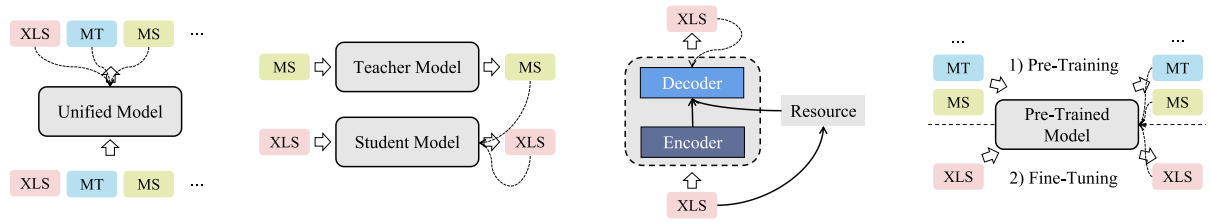
- When using extractive or compressive summarizers, the summarizers of the Trans-Sum methods can benefit from bilingual documents while the counterpart of the Sum-Trans methods can only utilize the source-language documents. Thus, the Trans-Sum methods typically achieve better performance than the Sum-Trans counterparts. For instance, on the manually translated DUC 2001 dataset, PBCS (Yao et al., 2015), as a Trans-Sum method, outperforms its Sum-Trans baseline by 8%/8.4%/10.4% in terms of ROUGE-1/2/L. On the other hand, the Trans-Sum methods are less efficient since they need to translate the whole documents rather than a few summaries.

- Apart from the above discussion, when adopting abstractive summarizers, a large-scale MS dataset is required to train the summarizers. It is also worth noting that the MS datasets in low-resource languages are much smaller than the MT counterparts (Tiedemann and Thottingal, 2020; Hasan et al., 2021b). Thus, the Trans-Sum methods are helpful if the source language is low-resource. In contrast, if the target language is low-resource in MS, the Sum-Trans methods are more useful (Ouyang et al., 2019; Ladhak et al., 2020).

5 End-to-End Methods

Though the pipeline method is intuitive, it 1) suffers from error propagation; 2) needs either a large corpus to train MT models or the monetary cost of paid MT services; 3) has a latency during inference. Thanks to the rapid development of neural networks, many end-to-end XLS models are proposed to alleviate the above issues.

In this section, we take stock of previous end-to-end XLS models and further divide them into four frameworks (cf., Figure 1): multi-task framework (§ 5.1), knowledge-distillation framework, (§ 5.2), resource-enhanced framework (§ 5.3), and pre-training framework (§ 5.4). For each framework, we will entirely introduce its core idea and



(a) Multi-Task Framework (b) Knowledge-Distillation Framework (c) Resource-Enhanced Framework (d) Pre-Training Framework

Figure 1: Overview of four end-to-end frameworks (best viewed in color). XLS: cross-lingual summarization; MT: machine translation; MS: monolingual summarization. Dashed arrows indicate the supervised signals. Rimless colored blocks denote the input or output sequences of the corresponding tasks. Note that the knowledge-distillation framework might contain more than one teacher model, and the auxiliary/pre-training tasks used in the multi-task/pre-training framework are not limited to MT and MS, here we omit these for simplicity.

corresponding models. Lastly, we discuss the pros and cons with respect to each framework (§ 5.5).

5.1 Multi-Task Framework

It is challenging for an end-to-end model to directly conduct XLS since it requires both the abilities to translate and summarize (Cao et al., 2020). As shown in Figure 1(a), many researchers use the related tasks (e.g., MT and MS) together with XLS to train unified models. In this way, XLS models could also benefit from the related tasks.

Zhu et al. (2019) utilize a shared transformer encoder to encode the input sequences of both XLS and MT/MS. Then, two independent transformer decoders are used to conduct XLS and MT/MS, respectively. This was the first paper to show that the end-to-end method outperforms the pipeline ones. Later, Cao et al. (2020) use two encoder-decoder models to perform MS in the source and target languages, respectively. Meanwhile, the source encoder and the target decoder jointly conduct XLS. Then, two linear mappers are used to convert the context representation (i.e., the output of encoders) from the source to the target language and vice versa. In addition, two discriminators are adopted to discriminate between the encoded and mapped representations. Thereby, the overall model could jointly learn to summarize documents and align representations between both languages.

Although the above efforts design unified models in the multi-task framework, their decoders are independent for different tasks, leading to limitations in capturing the relationships among the multiple tasks. To solve this problem, Takase and Okazaki (2020) train a single encoder-decoder model on both MS, MT and XLS datasets. They prepend a special token at the input sequences to indicate which task is performed. In addition, Bai et al. (2021a) make the MS a prerequisite for XLS and propose MCLAS, a XLS model of single encoder-decoder architecture. For the given documents, MCLAS generates the sequential concatenation of the corresponding monolingual and cross-lingual summaries. In this way, the translation alignment is also implicit in the generation process, making MCLAS achieve great performance in XLS. More recently, Liang et al. (2022) utilize conditional variational auto-encoder (CVAE) (Sohn et al., 2015) to capture the hierarchical relationship among MT, MS, and XLS. Specifically, three variables are adopted in the proposed model to reconstruct the results of MT, MS, and XLS, respectively. Besides, the used encoder and decoder are shared among all tasks, while the prior and recognition networks are independent to indicate the different tasks. Considering the limited XLS data in low-resource languages, Bai et al. (2021a) and Liang et al. (2022) also investigate XLS in the few-shot setting.

5.2 Knowledge-Distillation Framework

The original thought of knowledge distillation is distilling the knowledge in an ensemble of models (i.e., teacher models) into a single model (i.e., student model) (Hinton et al., 2015). Due to the close relationship between MT/MS and XLS, some researchers attempt to use MS or MT, or both models to teach the XLS model in the knowledge-distillation framework. In this way, besides the XLS labels, the student model can also learn from the output or hidden state of the teacher models.

Ayana et al. (2018) utilize large-scale MS and MT corpora to train MS and MT models, respectively. Then, they use the trained MS or MT, or both models, as the teacher models to teach the XLS student model. Both the teacher and student models are bi-directional GRU models (Cho et al., 2014). To let the student model mimic the output of the teacher model, the KL-divergence between the generation probabilities of these two models is used as the training objective. Later, Duan et al. (2019) implement transformer (Vaswani et al., 2017) as the backbone of the MS teacher model and the XLS student model, and further train the student model with two objectives: (1) the cross-entropy between the generation distributions of these two models; (2) the Euclidean distance between the attention weights of both models. It is worth noting that both Ayana et al. (2018) and Duan et al. (2019) focus on zero-shot XLS due to the scarcity of XLS datasets at that time, while their training objectives do not include XLS.

After the emergence of large-scale XLS datasets, Nguyen and Luu (2022) confirm that the knowledge-distillation framework can also be adopted in rich-resource scenarios. Specifically, they employ the transformer student and teacher models, and further propose a variant Sinkhorn divergence, which together with the XLS objective supervises the student XLS model.

5.3 Resource-Enhanced Framework

As shown in Figure 1(c), the resource-enhanced framework utilizes additional resources to enrich the information of the input documents, and the generation probability of the output summaries is conditioned on both the encoded and enriched information.

Zhu et al. (2020) explore the translation pattern in XLS. In detail, they first encode the input documents in source language via a transformer encoder, and then obtain the translation distribution for the words of the input documents by the `fast-align` toolkit (Dyer et al., 2013). Lastly, a transformer decoder is used to generate summaries in target language based on both its output distribution and the translation distributions. In this way, the extra bilingual alignment information helps the XLS model better learn the transformation from the source to the target language. Jiang et al. (2022) utilize the `TextRank` toolkit (Mihalcea and Tarau, 2004) to extract key clues from input sequences, and then construct article graphs based on these clues via a designed algorithm. Next, they encode the clues and the article graphs by a clue encoder (with transformer encoder architecture) and a graph encoder (based on graph nerual networks), respectively. Finally, a transformer decoder with two types of cross-attention (performed on the outputs of both clue and graph encoders) is adopted to generate final summaries. In addition, they consider the translation distribution used in Zhu et al. (2020) to further strength the proposed model.

5.4 Pre-Training Framework

The emergence of pre-trained models has brought NLP to a new era (Qiu et al., 2020). The pretrained models typically first learn the general representation from large-scale corpora, and then adapt to the specific task through fine-tuning.

More recently, the general multi-lingual pre-trained generative models have shown impressive performance on many multi-lingual NLP tasks. For example, mBART (Liu et al., 2020), as a multi-lingual pre-trained model, is derived from BART (Lewis et al., 2020). mBART is pre-trained with BART-style denoising objectives on a huge volume of unlabeled multi-lingual data. mBART shows its superiority in MT originally (Liu et al., 2020), and Liang et al. (2022) find it can also outperform many multi-task XLS models on large-scale XLS datasets through simply fine-tuning. Later, mBART-50 (Tang et al., 2021) goes a step further and extends the language processing abilities of mBART from 25 languages to 50 languages. In addition to the BART-style pre-trained models, mT5 (Xue et al., 2021) is a multi-lingual T5 (Raffel et al., 2020) model, which is pre-trained in 101 languages with the T5-style span corruption objective. Although great performance has been achieved, these general pre-trained models only utilize the denoising or span corruption objectives in multiple languages without any cross-lingual supervision, resulting in the under-explored cross-lingual ability.

To solve this problem, Xu et al. (2020) propose a mix-lingual XLS model which is pre-trained with masked language model (MLM), denoising auto-encoder (DAE), MS, translation span corruption (`TSC`), and `MT` tasks.[16] The `TSC` and `MT`

[16]Typewriter font indicates the cross-lingual tasks.

Pre-Training Task	Inputs	Targets
Machine Translation (MT)	Everything that kills make me feel alive	沉舟侧畔千帆过，病树前头万木春
Cross-Lingual Summarization (XLS)	Everything that kills make me feel alive	向死而生
Translation Span Corruption (TSC)	Everything [M1] make me [M2] alive. 沉舟侧畔千帆过，病树前头万木春。	[M1] that kills [M2] feel
Translation Pair Span Corruption (TPSC)	Everything that [M1] me feel alive. 沉舟侧畔[M2]过，病[M3]前头万木春。	[M1] kills make [M2] 千帆 [M3] 树

Table 3: Examples of inputs and targets used by different cross-lingual pre-training tasks for the sentence ''Everything that kills make me feel alive'' with its Chinese translation and summarization. The randomly selected spans are replaced with unique mask tokens (i.e, [M1], [M2], and [M3]) in TSC and TPSC.

pre-training samples are derived from OPUS English↔Chinese parallel corpus.[17] Dou et al. (2020) utilize XLS, MT and MS tasks to pre-train another XLS model. They leverage the English↔German/Chinese MT samples from the WMT2014/WMT2017 dataset. For XLS, they pre-train the model on En2ZhSum and English-German datasets (Dou et al., 2020). Wang et al. (2022b) focus on dialogue-oriented XLS and extend mBART-50 with MS, MT, and two dialogue-oriented pre-training objectives (i.e., action infilling and utterance permutation) via the second pre-training stage on MediaSum and XMediaSum datasets. Note that Xu et al. (2020), Dou et al. (2020), and Wang et al. (2022b) only focus on the XLS task. The languages supported by these models are limited to a few specific ones.

Furthermore, mT6 (Chi et al., 2021a) and ΔLM (Ma et al., 2021) are presented towards general cross-lingual abilities. In detail, Chi et al. (2021a) first present three tasks, namely, MT, TSC, and translation pair span corruption (TPSC), to extend mT5, and then design a PNAT decoding strategy to let the model separately decode each target span of SC-like pre-training tasks. Finally, Chi et al. (2021a) combine SC, TSC, and PNAT to jointly train the mT6 model. To support multiple languages, mT6 is pre-trained on CC-Net (Wenzek et al., 2020), MultiUN (Ziemski et al., 2016), IIT Bombay (Kunchukuttan et al., 2018), OPUS, and WikiMatrix (Schwenk et al., 2021) corpora, covering a total of 94 languages. ΔLM reuses the parameters of InfoXLM (Chi et al., 2021b) and further is trained with SC and TSC tasks on CC100 (Conneau et al., 2020), CC-Net, Wikipedia dump, CCAligned (El-Kishky et al., 2020), and OPUS corpora, including 100 languages. The superiority of mT6 and ΔLM on WikiLingua (a large-scale XLS dataset) has been demonstrated. Moreover, there are also some general cross-lingual pre-trained models that have not been evaluated in XLS, for example, XNLG (Chi et al., 2020) and VECO (Luo et al., 2021).

Table 3 shows the details of the above cross-lingual pre-training tasks. TSC and TPSC predict the masked spans from a translation pair. The input sequence of TSC is only masked in one language while the counterpart of TPSC is masked in both languages.

5.5 Discussion

Table 4 summarizes all end-to-end XLS models. We conclude that all four frameworks resort to external resources to improve XLS performance: (1) The multi-task framework uses large-scale MS and MT corpora to help XLS. Though the multi-task learning is intuitive, its training strategy and weights of different task is non-trivial to determine. (2) The knowledge-distillation framework is another way to utilize the large-scale MS and MT corpora. This framework is most suitable for zero-shot XLS since it could be supervised by the MS and MT teacher models without any XLS labels. Nevertheless, knowledge distillation often fails to live up to its name, transferring very limited knowledge from teacher to student (Stanton et al., 2021). Thus, it should be verified more deeply in the rich-resource XLS. (3) The resource-enhanced framework employs the off-the-shelf toolkits to enhance the representation of input documents. This framework significantly relaxes the dependence on external data, but it suffers from error propagation. (4) The pre-training framework can benefit from both unlabeled and labeled corpora. In detail, pre-trained models learn

[17] http://opus.nlpl.eu/.

Model	Architecture	Training Objective	Evaluation Direction	Evaluation Dataset
		Multi-Task Framework		
CLS+MS (Zhu et al., 2019)	Transformer	XLS+MS	En↔Zh	En2ZhSum, Zh2EnSum
CLS+MT (Zhu et al., 2019)	Transformer	XLS+MT	En↔Zh	En2ZhSum, Zh2EnSum
Cao et al. (2020)	Transformer	XLS+MS+REC	En↔Zh	Gigaword†, DUC2004†, En2ZhSum, Zh2EnSum
Transum (Takase and Okazaki, 2020)	Transformer	XLS+MS+MT	Ar/Zh→En, En→Ja	DUC2004†, JAMUL†
MCLAS (Bai et al., 2021a)	Transformer	XLS+MS	En↔Zh, En→De	En2ZhSum, Zh2EnSum, En2DeSum
VHM (Liang et al., 2022)	Transformer*	XLS+MS+MT	En↔Zh	En2ZhSum, Zh2EnSum
		Knowledge-Distillation Framework		
MS teacher (Ayana et al., 2018)	GRU	XLS+*KD* (MS)	En→Zh	DUC2003†, DUC2004†
MT teacher (Ayana et al., 2018)	GRU	XLS+*KD* (MT)	En→Zh	DUC2003†, DUC2004†
MS+MT teachers (Ayana et al., 2018)	GRU	XLS+*KD* (MS+MT)	En→Zh	DUC2003†, DUC2004†
Duan et al. (2019)	Transformer	XLS+*KD* (MS)	Zh→En	Gigaword†, DUC2004†
Nguyen and Luu (2022)	Transformer	XLS+*KD* (MS)	En↔Zh, En↔Ja, En→Ar/Vi	En2ZhSum, Zh2EnSum, WikiLingua
		Resource-Enhanced Framework		
ATS (Zhu et al., 2020)	Transformer*	XLS	En↔Zh	En2ZhSum, Zh2EnSum
GlueGraphSum (Jiang et al., 2022)	Transformer*	XLS	En↔Zh	En2ZhSum, Zh2EnSum, CyEn2ZhSum‡
		Pre-Training Framework		
Xu et al. (2020)	Transformer	MLM+DAE+MS+MT+TSC	En↔Zh	En2ZhSum, Zh2EnSum
Dou et al. (2020)	Transformer	XLS+MT+MS	En→Zh, En→De	En2ZhSum, English-German‡
mT6 (Chi et al., 2021a)	Transformer	SC+TSC+PNAT	Es/Ru/Tr/Vi→En	WikiLingua
ΔLM (Ma et al., 2021)	Transformer	SC+TSC	Es/Ru/Tr/Vi→En	WikiLingua
mDIALBART (Wang et al., 2022b)	Transformer	AcI+UP+MS+MT	En→Zh, En→De	XMediaSum40k

Table 4: The summary of end-to-end XLS models. ''*Transformer*'' means the vanilla transformer encoder-decoder architecture. $*$ denotes the variant architecture. ''REC'' represents the reconstruction objective, which is used to supervise the linear mappers in the model proposed by Cao et al. (2020). ''*KD*'' denotes the knowledge distillation objectives, derived from the output or hidden state of the corresponding teacher models, such as MS and MT models. The ''*Training Objective*'' of pre-trained models lists the pre-training objectives. Language nomenclature used in ''*Evaluation Direction*'' is ISO 639-1 codes. † indicates the number of samples in the dataset is less than 2000. ‡ denotes unreleased datasets.

Model	En2ZhSum			Zh2EnSum		
	R-1	R-2	R-L	R-1	R-2	R-L
CLS+MS♡† (Zhu et al., 2019)	38.25	20.20	34.76	40.34	22.65	36.39
CLS+MT♡† (Zhu et al., 2019)	40.23	22.32	36.59	40.25	22.58	36.21
Cao et al. (2020)♡†	38.12	16.76	33.86	40.97	23.20	36.96
VHM♡* (Liang et al., 2022)	40.98	23.07	37.12	41.36	24.64	37.15
ATS (Zhu et al., 2020)♣†	40.47	22.21	36.89	40.68	24.12	36.97
mBART (Liu et al., 2020)♠‡	41.55	23.27	37.22	43.61	25.14	38.79
Dou et al. (2020)♠*	42.83	23.30	**39.29**	–	–	–
Xu et al. (2020)♠*	**43.50**	**25.41**	29.66	41.62	23.35	37.26
mVHM (Liang et al., 2022)♡♠*	41.95	23.54	37.67	**43.97**	**25.61**	**39.19**

Table 5: The leaderboard of end-to-end XLS models on En2ZhSum and Zh2EnSum datasets (Zhu et al., 2019) in terms of ROUGE(R)-1/2/L (Lin, 2004). The evaluation scripts refer to Zhu et al. (2020). ♡: multi-task framework; ♣: resource-enhanced framework; ♠: pre-training framework. † indicates the results are obtained by evaluating output files provided by the authors; ‡ denotes the results by running their released codes; * indicates the results are reported in the original papers which adopt the same evaluation scripts as Zhu et al. (2020).

the general language knowledge from large-scale unlabeled data with self-supervised objectives. In order to improve the cross-lingual ability, they can resort to MT parallel corpus and design supervised signals. This framework absorbs more knowledge from more external corpora than others, leading to the promising performance on XLS.

To give a deeper comparison of end-to-end XLS models, as shown in Table 5, we organize a leaderboard with unified evaluation metrics, based on the released code and generated results from representative published literature. The models in the pre-training framework (Liu et al., 2020; Dou et al., 2020; Xu et al., 2020) generally outperform others. Additionally, the pre-training framework could also serve other frameworks. For example, Liang et al. (2022) utilize mBART weights as model initialization for VHM (i.e., mVHM), bringing decent gains compared with vanilla VHM. Therefore, it is possible and valuable to combine the advantages of different frameworks, which is worthy of discussion in the future.

6 Prospects

In this section, we discuss and suggest the following promising future directions, which meet actual application needs:

The Essence of XLS. Unifying two abilities (i.e., translation and summarization abilities) in

a single model is non-trivial (Cao et al., 2020). Even though the effectiveness of the state-of-the-art models has been proved, the essence of XLS remains unclear, especially (1) the hierarchical relationship between MT&MS and XLS (Liang et al., 2022), and (2) the theoretical analysis for *what makes MT&MS help XLS?*

XLS Dataset with Low-Resource Languages. There are thousands of languages in the world and most of them are low-resource. Despite the practical significance, building high-quality and large-scale XLS datasets whose source or target language is low-resource remains challenging (c.f., Section 3.3), and needs to be further explored in the future.

Unified XLS across Genres and Domains. As we described in Section 3, existing XLS datasets cover multiple genres or domains, namely, news, dialogue, guides, and encyclopedia. The diversity across them is naturally promoting the need for unified XLS, instead of promoting the trend of devising unique models on individual genres or domains. At present, the unified XLS is still under-explored, making us believe the urgent need for it.

Controllable XLS. Bai et al. (2021b) integrate a compression rate to control how much information should be kept in the target language. If the compression rate is 100%, XLS degrades to MT. Thus, the continuous variable unifies XLS and MT tasks. In this manner, a new research view is introduced to leverage MT to help XLS. In addition, controlling some other attributes of the target summary may be useful in real applications, such as entity-centric XLS and aspect-based XLS.

Low-Resource XLS. Most languages in the world are low-resource, which makes large-scale parallel datasets across these languages rare and expensive. Hence, low-resource XLS is more realistic. Nevertheless, current work has not well investigate and explore this situation. Recently, prompt-based learning has become a new paradigm in NLP (Liu et al., 2021). With the help of the well-designed prompting function, a pre-trained model is able to perform few-shot or even zero-shot learning. Future work can adopt prompt-based learning to deal with the low-resource XLS.

Triangular XLS. Following triangular MT, triangular XLS is a special case of low-resource XLS where the language pair of interest has limited parallel data, but both languages have abundant parallel data with a pivot language. This situation typically appears in multi-lingual website datasets (a category of XLS datasets, cf., § 3.2), because their documents are usually centered in English and then translated into other languages to facilitate global users. Hence, English acts as the pivot language. How to exploit such abundant parallel data to improve the XLS of interest language pairs remains challenging.

Many-to-Many XLS. Most previous work trains XLS models separately in each cross-lingual direction. In this way, the knowledge of XLS cannot be transferred among different directions. Besides, a trained model can only perform in a single direction, resulting in limited usage. To solve this problem, Hasan et al. (2021a) jointly fine-tune mT5 in multiple directions. Lastly, the fine-tuned model can perform in arbitrary even unseen directions, which is named many-to-many XLS. Future work can focus on designing robust and effective training strategies for many-to-many XLS.

Long Document XLS. Recently, long document MS has attracted wide research attention (Cohan et al., 2018; Sharma et al., 2019; Wang et al., 2021, 2022a). Long document XLS is also important in real scenes, for example, facilitating researchers to access the arguments of scientific papers in foreign languages. Nevertheless, this direction has not been noticed by previous work. Interestingly, we find many non-English scientific papers have corresponding English abstracts due to the regulations of publishers. For example, many Chinese academic journals require researchers to write abstracts in both Chinese and English. This might be a feasible method to construct long document XLS datasets. We hope future work can promote this direction.

Multi-Document XLS. Previous multi-document XLS work (Orăsan and Chiorean, 2008; Boudin et al., 2011; Zhang et al., 2016) only utilizes statistical features to build pipeline systems, and further performs on early XLS datasets. The multi-document XLS is also worthy of discussion in the pre-trained model era.

Multi-Modal XLS. With the increasing of multimedia data on the internet, some researchers have put their effort into multi-modal summarization (Zhu et al., 2018; Sanabria et al., 2018; Li et al., 2018, 2020; Fu et al., 2021), where the input of summarization systems is a document together with images or videos. Nevertheless, existing multi-modal summarization work only focuses on the monolingual scene and ignores cross-lingual ones. We hope future work could highlight multi-modal XLS.

Evaluation Metrics. Developing evaluation metrics for XLS is still an open problem that needs to be further studied. Current XLS metrics typically inherit from MS. However, different from MS, the XLS samples consist of ⟨source document, (target document), source summary, target summary⟩. Besides the target summary, how to apply other information to assess the summary quality would be an interesting point for further study.

Others. Considering that current XLS research is still in the preliminary stage, many research points of MS are missing in XLS, such as the factual inconsistency and hallucination problems. These directions are also worthy to be deeply explored in further work.

7 Conclusion

In this paper, we present the first comprehensive survey of current research efforts on XLS. We systematically summarize existing XLS datasets and methods, highlight their characteristics, and compare them with each other to provide deeper analyses. In addition, we give multiple perspective directions to facilitate further research on XLS. We hope that this XLS survey can provide a clear picture of this topic and boost the development of the current XLS technologies.

Acknowledgments

We would like to thank anonymous reviewers for their suggestions and comments. This research is supported by the National Key Research and Development Project (no. 2020AAA0109302), the National Natural Science Foundation of China (no. 62072323, 62102276), Shanghai Science and Technology Innovation Action Plan (no. 19-511120400), Shanghai Municipal Science and Technology Major Project (no. 2021SHZDZX01-03), the Natural Science Foundation of Jiangsu Province (grant no. BK20210705), the Natural Science Foundation of Educational Commission of Jiangsu Province, China (grant no. 21KJD52-0005), and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

Ayana, Shi-qi Shen, Yun Chen, Cheng Yang, Zhi-yuan Liu, and Mao-song Sun. 2018. Zero-shot cross-lingual neural headline generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(12):2319–2327. https://doi.org/10.1109/TASLP.2018.2842432

Yu Bai, Yang Gao, and Heyan Huang. 2021a. Cross-lingual abstractive summarization with limited parallel resources. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6910–6924, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.acl-long.538

Yu Bai, Heyan Huang, Kai Fan, Yang Gao, Zewen Chi, and Boxing Chen. 2021b. Bridging the gap: Cross-lingual summarization with compression rate. *ArXiv preprint*, abs/2110.07936v1.

Florian Boudin, Stéphane Huet, and Juan-Manuel Torres-Moreno. 2011. A graph-based approach to cross-language multi-document summarization. *Polibits*, 43:113–118.

Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.

Yue Cao, Hui Liu, and Xiaojun Wan. 2020. Jointly learning to align and summarize for neural cross-lingual summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6220–6231, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.554

Zewen Chi, Li Dong, Shuming Ma, Shaohan Huang, Saksham Singhal, Xian-Ling Mao, Heyan Huang, Xia Song, and Furu Wei. 2021a. mT6: Multilingual pretrained text-to-text transformer with translation pairs. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1671–1683, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.125

Zewen Chi, Li Dong, Furu Wei, Wenhui Wang, Xian-Ling Mao, and Heyan Huang. 2020. Cross-lingual natural language generation via pre-training. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7–12, 2020*, pages 7570–7577. AAAI Press. https://doi.org/10.1609/aaai.v34i05.6256

Zewen Chi, Li Dong, Furu Wei, Nan Yang, Saksham Singhal, Wenhui Wang, Xia Song, Xian-Ling Mao, Heyan Huang, and Ming Zhou. 2021b. InfoXLM: An information-theoretic framework for cross-lingual language model pre-training. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3576–3588, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.280

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar. Association for Computational Linguistics.

Arman Cohan, Franck Dernoncourt, Doo Soon Kim, Trung Bui, Seokhwan Kim, Walter Chang, and Nazli Goharian. 2018. A discourse-aware attention model for abstractive summarization of long documents. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 615–621, New Orleans, Louisiana. Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-2097

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.747

Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning*, 20(3):273–297. https://doi.org/10.1007/BF00994018

Zi-Yi Dou, Sachin Kumar, and Yulia Tsvetkov. 2020. A deep reinforced model for zero-shot cross-lingual summarization with bilingual semantic similarity rewards. In *Proceedings of the Fourth Workshop on Neural Generation and Translation*, pages 60–68, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.ngt-1.7

Xiangyu Duan, Mingming Yin, Min Zhang, Boxing Chen, and Weihua Luo. 2019. Zero-shot cross-lingual abstractive sentence summarization through teaching generation and attention. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3162–3172, Florence, Italy. Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1305

Chris Dyer, Victor Chahuneau, and Noah A. Smith. 2013. A simple, fast, and effective reparameterization of IBM model 2. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–648, Atlanta, Georgia. Association for Computational Linguistics.

Ahmed El-Kishky, Vishrav Chaudhary, Francisco Guzmán, and Philipp Koehn. 2020. CCAligned: A massive collection of cross-lingual web-document pairs. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5960–5969, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.480

Mehwish Fatima and Michael Strube. 2021. A novel Wikipedia based dataset for monolingual and cross-lingual summarization. In *Proceedings of the Third Workshop on New Frontiers in Summarization*, pages 39–50, Online and in Dominican Republic. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.newsum-1.5

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022a. Language-agnostic BERT sentence embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 878–891, Dublin, Ireland. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.62

Xiachong Feng, Xiaocheng Feng, and Bing Qin. 2022b. MSAMSum: Towards benchmarking multi-lingual dialogue summarization. In *Proceedings of the Second DialDoc Workshop on Document-grounded Dialogue and Conversational Question Answering*, pages 1–12, Dublin, Ireland. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.dialdoc-1.1

Xiachong Feng, Xiaocheng Feng, and Bing Qin. 2022c. A survey on dialogue summarization: Recent advances and new frontiers. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 5453–5460. International Joint Conferences on Artificial Intelligence Organization. https://doi.org/10.24963/ijcai.2022/764

Xiyan Fu, Jun Wang, and Zhenglu Yang. 2021. MM-AVS: A full-scale dataset for multi-modal summarization. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5922–5926, Online. Association for Computational Linguistics.

George Giannakopoulos. 2013. Multi-document multilingual summarization and evaluation tracks in ACL 2013 MultiLing workshop. In *Proceedings of the MultiLing 2013 Workshop on Multilingual Multi-document Summarization*, pages 20–28, Sofia, Bulgaria. Association for Computational Linguistics.

George Giannakopoulos, Jeff Kubina, John Conroy, Josef Steinberger, Benoit Favre, Mijail Kabadjov, Udo Kruschwitz, and Massimo Poesio. 2015. MultiLing 2015: Multilingual summarization of single and multi-documents, on-line fora, and call-center conversations. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 270–274, Prague, Czech Republic. Association for Computational Linguistics. https://doi.org/10.18653/v1/W15-4638

Bogdan Gliwa, Iwona Mochol, Maciej Biesek, and Aleksander Wawer. 2019. SAMSum corpus: A human-annotated dialogue dataset for abstractive summarization. In *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, pages 70–79, Hong Kong, China. Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-5409

Tahmid Hasan, Abhik Bhattacharjee, Wasi Uddin Ahmad, Yuan-Fang Li, Yong-Bin Kang, and Rifat Shahriyar. 2021a. Crosssum: Beyond English-centric cross-lingual abstractive text summarization for 1500+ language pairs. *ArXiv preprint*, abs/2112.08804v1.

Tahmid Hasan, Abhik Bhattacharjee, Md. Saiful Islam, Kazi Mubasshir, Yuan-Fang Li, Yong-Bin Kang, M. Sohel Rahman, and Rifat Shahriyar. 2021b. XL-sum: Large-scale multilingual abstractive summarization for 44 languages. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4693–4703, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.findings-acl.413

Karl Moritz Hermann, Tomás Kociský, Edward Grefenstette, Lasse Espeholt, Will Kay,

Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 1693–1701.

Geoffrey E. Hinton, Oriol Vinyals, and Jeffrey Dean. 2015. Distilling the knowledge in a neural network. *ArXiv preprint*, abs/1503.02531v1.

Baotian Hu, Qingcai Chen, and Fangze Zhu. 2015. LCSTS: A large scale Chinese short text summarization dataset. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1967–1972, Lisbon, Portugal. Association for Computational Linguistics.

Shuyu Jiang, Dengbiao Tu, Xingshu Chen, R. Tang, Wenxian Wang, and Haizhou Wang. 2022. ClueGraphSum: Let key clues guide the cross-lingual abstractive summarization. *ArXiv preprint*, abs/2203.02797v2.

Fajri Koto, Jey Han Lau, and Timothy Baldwin. 2021. Evaluating the efficacy of summarization evaluation across languages. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 801–812, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.findings-acl.71

Anoop Kunchukuttan, Pratik Mehta, and Pushpak Bhattacharyya. 2018. The IIT Bombay English-Hindi parallel corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Faisal Ladhak, Esin Durmus, Claire Cardie, and Kathleen McKeown. 2020. WikiLingua: A new benchmark dataset for cross-lingual abstractive summarization. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4034–4048, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.findings-emnlp.360

Anton Leuski, Chin-Yew Lin, Liang Zhou, Ulrich Germann, Franz Josef Och, and Eduard H. Hovy. 2003. Cross-lingual c*st*rd: English access to hindi information. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2:245–269. https://doi.org/10.1145/979872.979877

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.703

Haoran Li, Junnan Zhu, Tianshang Liu, Jiajun Zhang, and Chengqing Zong. 2018. Multimodal sentence summarization with modality attention and image filtering. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 4152–4158. International Joint Conferences on Artificial Intelligence Organization. https://doi.org/10.24963/ijcai.2018/577

Mingzhe Li, Xiuying Chen, Shen Gao, Zhangming Chan, Dongyan Zhao, and Rui Yan. 2020. VMSMO: Learning to generate multimodal summary for video-based news articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9360–9369, Online. Association for Computational Linguistics.

Yunlong Liang, Fandong Meng, Chulun Zhou, Jinan Xu, Yufeng Chen, Jinsong Su, and Jie Zhou. 2022. A variational hierarchical model for neural cross-lingual summarization. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2088–2099, Dublin, Ireland. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.148

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81,

Barcelona, Spain. Association for Computational Linguistics.

Elvys Linhares Pontes, Stéphane Huet, Juan-Manuel Torres-Moreno, and Andréa Carneiro Linhares. 2018. Cross-language text summarization using sentence and multi-sentence compression. In *Natural Language Processing and Information Systems*, pages 467–479, Cham. Springer International Publishing.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2021. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ArXiv preprint*, abs/2107.13586v1.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.

Fuli Luo, Wei Wang, Jiahao Liu, Yijia Liu, Bin Bi, Songfang Huang, Fei Huang, and Luo Si. 2021. VECO: Variable and flexible cross-lingual pre-training for language understanding and generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3980–3994, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.acl-long.308

Shuming Ma, Li Dong, Shaohan Huang, Dongdong Zhang, Alexandre Muzio, Saksham Singhal, Hany Hassan Awadalla, Xia Song, and Furu Wei. 2021. DeltaLM: Encoder-decoder pre-training for language generation and translation by augmenting pretrained multilingual encoders. *ArXiv preprint*, abs/2106.13736v2.

Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 404–411, Barcelona, Spain. Association for Computational Linguistics.

Nathan Ng, Kyra Yee, Alexei Baevski, Myle Ott, Michael Auli, and Sergey Edunov. 2019. Facebook FAIR's WMT19 news translation task submission. In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 314–319, Florence, Italy. Association for Computational Linguistics.

Khanh Nguyen and Hal Daumé III. 2019. Global Voices: Crossing borders in automatic news summarization. In *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, pages 90–97, Hong Kong, China. Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-5411

Thong Thanh Nguyen and Anh Tuan Luu. 2022. Improving neural cross-lingual abstractive summarization via employing optimal transport distance for knowledge distillation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(10):11103–11111. https://doi.org/10.1609/aaai.v36i10.21359

Constantin Orăsan and Oana Andreea Chiorean. 2008. Evaluation of a cross-lingual Romanian-English multi-document summarizer. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Jessica Ouyang, Boya Song, and Kathy McKeown. 2019. A robust abstractive system for cross-lingual summarization. In *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)*, pages 2025–2031, Minneapolis, Minnesota. Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1204

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab.

Chris D. Paice. 1990. Constructing literature abstracts by computer: Techniques and prospects. *Information Processing & Management*, 26(1):171–186. https://doi.org/10.1016/0306-4573(90)90014-S

Laura Perez-Beltrachini and Mirella Lapata. 2021. Models and datasets for cross-lingual summarisation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9408–9423, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.742

Xipeng Qiu, Tianxiang Sun, Yige Xu, Yunfan Shao, Ning Dai, and Xuanjing Huang. 2020. Pre-trained models for natural language processing: A survey. *ArXiv preprint*, abs/2003.08271v4.

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

Ramon Sanabria, Ozan Caglayan, Shruti Palaskar, Desmond Elliott, Loïc Barrault, Lucia Specia, and Florian Metze. 2018. How2: A large-scale dataset for multimodal language understanding. In *Proceedings of the Workshop on Visually Grounded Interaction and Language (ViGIL)*. NeurIPS.

Holger Schwenk, Vishrav Chaudhary, Shuo Sun, Hongyu Gong, and Francisco Guzmán. 2021. WikiMatrix: Mining 135M parallel sentences in 1620 language pairs from Wikipedia. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1351–1361, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.eacl-main.115

Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083, Vancouver, Canada. Association for Computational Linguistics.

Eva Sharma, Chen Li, and Lu Wang. 2019. BIGPATENT: A large-scale dataset for abstractive and coherent summarization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2204–2213, Florence, Italy. Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1212

Kihyuk Sohn, Honglak Lee, and Xinchen Yan. 2015. Learning structured output representation using deep conditional generative models. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7–12, 2015, Montreal, Quebec, Canada*, pages 3483–3491.

Samuel Stanton, Pavel Izmailov, Polina Kirichenko, Alexander A. Alemi, and Andrew G. Wilson. 2021. Does knowledge distillation really work? *Advances in Neural Information Processing Systems*, 34.

Sho Takase and Naoaki Okazaki. 2020. Multi-task learning for cross-lingual abstractive summarization. *ArXiv preprint*, abs/2010.07503v1.

Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, Jiatao Gu, and Angela Fan. 2021. Multilingual translation from denoising pre-training. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3450–3466, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.findings-acl.304

Jörg Tiedemann and Santhosh Thottingal. 2020. OPUS-MT – building open translation services for the world. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 479–480, Lisboa, Portugal. European Association for Machine Translation.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4–9, 2017, Long Beach, CA, USA*, pages 5998–6008.

Xiaojun Wan. 2011. Using bilingual information for cross-language document summarization.

In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 1546–1555, Portland, Oregon, USA. Association for Computational Linguistics.

Xiaojun Wan, Huiying Li, and Jianguo Xiao. 2010. Cross-language document summarization based on machine translation quality prediction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 917–926, Uppsala, Sweden. Association for Computational Linguistics.

Xiaojun Wan, Fuli Luo, Xue Sun, Songfang Huang, and Jin ge Yao. 2018. Cross-language document summarization via extraction and ranking of multiple summaries. *Knowledge and Information Systems*, 58:481–499.

Jiaan Wang, Zhixu Li, Qiang Yang, Jianfeng Qu, Zhigang Chen, Qingsheng Liu, and Guoping Hu. 2021. Sportssum2.0: Generating high-quality sports news from live text commentary. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 3463–3467, New York, NY, USA. Association for Computing Machinery. https://doi.org/10.1145/3459637.3482188

Jiaan Wang, Zhixu Li, Tingyi Zhang, Duo Zheng, Jianfeng Qu, An Liu, Lei Zhao, and Zhigang Chen. 2022a. Knowledge enhanced sports game summarization. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, WSDM '22, pages 1045–1053, New York, NY, USA. Association for Computing Machinery. https://doi.org/10.1145/3488560.3498405

Jiaan Wang, Fandong Meng, Ziyao Lu, Duo Zheng, Zhixu Li, Jianfeng Qu, and Jie Zhou. 2022b. ClidSum: A benchmark dataset for cross-lingual dialogue summarization. *ArXiv preprint*, abs/2202.05599v1.

Guillaume Wenzek, Marie-Anne Lachaux, Alexis Conneau, Vishrav Chaudhary, Francisco Guzmán, Armand Joulin, and Edouard Grave. 2020. CCNet: Extracting high quality monolingual datasets from web crawl data. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4003–4012, Marseille, France. European Language Resources Association.

Ruochen Xu, Chenguang Zhu, Yu Shi, Michael Zeng, and Xuedong Huang. 2020. Mixed-lingual pre-training for cross-lingual summarization. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 536–541, Suzhou, China. Association for Computational Linguistics.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.

Jin-ge Yao, Xiaojun Wan, and Jianguo Xiao. 2015. Phrase-based compressive cross-language summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 118–127. Association for Computational Linguistics. https://doi.org/10.18653/v1/D15-1012

Jiajun Zhang, Yu Zhou, and Chengqing Zong. 2016. Abstractive cross-language summarization via translation model enhanced predicate argument structure fusing. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24:1842–1853. https://doi.org/10.1109/TASLP.2016.2586608

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu. 2020a. PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. In *Proceedings of the 37th International Conference on Machine Learning, ICML* 2020, 13–18 July 2020, Virtual Event, volume 119 of *Proceedings of Machine Learning Research*, pages 11328–11339. PMLR.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020b.

Bertscore: Evaluating text generation with BERT. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26–30, 2020*. OpenReview.net.

Wei Zhao, Maxime Peyrard, Fei Liu, Yang Gao, Christian M. Meyer, and Steffen Eger. 2019. MoverScore: Text generation evaluating with contextualized embeddings and earth mover distance. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 563–578, Hong Kong, China. Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1053

Chenguang Zhu, Yang Liu, Jie Mei, and Michael Zeng. 2021. MediaSum: A large-scale media interview dataset for dialogue summarization. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5927–5934, Online. Association for Computational Linguistics.

Junnan Zhu, Haoran Li, Tianshang Liu, Yu Zhou, Jiajun Zhang, and Chengqing Zong. 2018. MSMO: Multimodal summarization with multimodal output. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4154–4164, Brussels, Belgium. Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1448

Junnan Zhu, Qian Wang, Yining Wang, Yu Zhou, Jiajun Zhang, Shaonan Wang, and Chengqing Zong. 2019. NCLS: Neural cross-lingual summarization. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3054–3064, Hong Kong, China. Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1302

Junnan Zhu, Yu Zhou, Jiajun Zhang, and Chengqing Zong. 2020. Attend, translate and summarize: An efficient method for neural cross-lingual summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1309–1321, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.121

Michał Ziemski, Marcin Junczys-Dowmunt, and Bruno Pouliquen. 2016. The United Nations parallel corpus v1.0. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3530–3534, Portorož, Slovenia. European Language Resources Association (ELRA).