

GreekBART: The First Pretrained Greek Sequence-to-Sequence Model

Iakovos Evdaimon*, Hadi Abdine*, Christos Xypolopoulos*[†],
Stamatis Outsios*, Michalis Vazirgiannis*[‡], Giorgos Stamou[†]

*École Polytechnique
Palaiseau, France

iakovos.evdaimon, hadi.abdine, christos.xypolopoulos@polytechnique.edu

[†] National Technical University of Athens
Athens, Greece

cxyolopoulos@mail.ntua.gr, gstam@cs.ntua.gr

* Athens University of Economics and Business
Athens, Greece
soutsios@aueb.gr

[‡] KTH Royal Institute of Technology
Stockholm, Sweden
mvaz@kth.se

Abstract

The era of transfer learning has revolutionized the fields of Computer Vision and Natural Language Processing, bringing powerful pretrained models with exceptional performance across a variety of tasks. Specifically, Natural Language Processing tasks have been dominated by transformer-based language models. In Natural Language Inference and Natural Language Generation tasks, the BERT model and its variants, as well as the GPT model and its successors, demonstrated exemplary performance. However, the majority of these models are pretrained and assessed primarily for the English language or on a multilingual corpus. In this paper, we introduce GreekBART, the first Seq2Seq model based on BART-base architecture and pretrained on a large-scale Greek corpus. We evaluate and compare GreekBART against BART-random, Greek-BERT, and XLM-R on a variety of discriminative tasks. In addition, we examine its performance on two NLG tasks from GreekSUM, a newly introduced summarization dataset for the Greek language. The model, the code, and the new summarization dataset will be publicly available.

Keywords: Natural language processing, Pre-training, Transfer learning, Sequence-to-sequence models, Natural Language Generation, Zero-shot learning.

1. Introduction & Related Work

Transfer learning has revolutionized the field of machine learning, particularly in Computer Vision and Natural Language Processing (Krizhevsky et al., 2017). It has become a popular approach to address the challenge of insufficient training data in real-world machine learning applications. Especially, transformer (Vaswani et al., 2017) based pretrained models, like BERT (Devlin et al., 2019) and its variants, are broadly used in Natural Language Processing, as have been shown to be effective in many tasks (Vaswani et al., 2017; Devlin et al., 2019).

BART (Lewis et al., 2020) is a denoising auto-encoder for pretraining sequence-to-sequence models. The unsupervised pretrained BART learns a language model that can be fine-tuned for specific NLP tasks. While BART is well-suited for machine translation, question-answering, and text summarization tasks, it is also effective in sequence clas-

sification tasks.

In recent years, extensive research has been done on various languages, excluding English resulting in language models like CamemBERT (Martin et al., 2020), BARThez (Kamal Eddine et al., 2021), CAMELBERT (Inoue et al., 2021), AraBART (Eddine et al., 2022), BETO (Cañete et al., 2020), NASes (Ahuir et al., 2021), and BARTpho (Tran et al., 2021). In addition to monolingual models, there has been a range of multilingual language models introduced with the objective of learning multiple languages concurrently by pretraining on diverse language corpora. M-BERT (Devlin et al., 2019), XLM (Conneau and Lample, 2019) and XLM-R (Conneau et al., 2020) are notable examples of such models, pretrained on a wide range of languages. Also, mBART (Liu et al., 2020), a multilingual version of BART, and its extension mBART50 (Tang et al., 2020) were pretrained on 25 and 50 languages, respectively. All the aforementioned multilingual models, except for mBART

and mBART50, support Greek. Recently, ChatGPT¹, based on GPT-3 (Brown et al., 2020) and GPT-4 (OpenAI, 2023) architectures, excels in generating high-quality text sequences for tasks like summarization and question answering. However, multilingual models cannot compete with the performance of monolingual models in most NLP tasks (Singh and Lefever, 2022).

Compared to widely spoken languages, Greek has limited linguistic resources and undeveloped research in deep learning models. However, efforts are underway to develop datasets, models, knowledge bases, and frameworks for Greek NLP. Outsios et al. (2018) presented the production of Greek word embeddings, where a large corpus of about 50GB, crawled from about 20 million URLs, was used for their work. Later, Lioudakis et al. (2020) presented an ensemble method, Continuous Bag-of-Skip-grams, for extracting word representations for Greek. Recently, Koutsikakis et al. (2020) employed Greek-BERT, the first transformer-based language model pretrained on a 29GB dataset, achieving state-of-the-art performance in Greek NLP tasks. It is worth noting that Papantoniou and Tzitzikas have provided a thorough survey (Papantoniou and Tzitzikas, 2020) of NLP work in Greek. In this study, we address the challenge posed by the limited effectiveness of multilingual models compared to monolingual models, as well as the scarcity of deep language models specifically designed for the Greek language. Thus, we propose GreekBART, the first pretrained Seq2Seq monolingual model for Greek. GreekBART is based on the BART-base architecture pretrained on a large monolingual Greek corpus. Our model exceeds the possibilities of Greek-BERT (Koutsikakis et al., 2020), focusing on generative tasks. GreekBART is evaluated on two different generative tasks and on four discriminative tasks. Our model² is openly accessible to those interested in extending its applications or benefiting from our contributions.

2. GreekBART

Our model is based on BART_{BASE} (Lewis et al., 2020). It consists of 6 encoder and 6 decoder layers with 768 hidden dimensions. Each encoder and decoder has 12 attention heads, and normalization layers are applied to both (Liu et al., 2020). The model has approximately 181M parameters. Our approach follows a similar methodology as Kamal Eddine et al. (2021), where a pretrained monolingual model is used in a different language from English, following the techniques of BART (Lewis et al., 2020) and mBART (Liu et al., 2020).

¹<https://openai.com/blog/chatgpt>

²<https://github.com/iakovosevdaimon/GreekBART>

Corpus	Size before deduplication	Size after deduplication
OSCAR	51.7	44.6
Greek Web Corpus	38.4	30.9
Wikipedia	0.9	0.9
EuroParl	0.5	0.5
Total	91.5	76.9

Table 1: Datasets which consists of the GreekBART pretraining corpus (sizes in GB, before and after cleaning and deduplication).

2.1. Pretraining corpus

The pretrained corpus for GreekBART consists of the Greek part of Wikipedia³, the Greek part of EuroParl (Koehn, 2005), the Greek part of OSCAR (a clean version of CommonCrawl) (Abadji et al., 2022), and the Greek Web Corpus (Outsios et al., 2018). These corpora cover a wide range of text types, including formal and informal text, news articles, encyclopedic information, and political conversations. The corpora were deduplicated using the runiq package⁴. The final corpus was tokenized using SentencePiece (Kudo and Richardson, 2018), resulting in a vocabulary of 50K sub-words and a character coverage of 99.95%. The corpus size was 76.9GB before tokenization and 87.6GB after tokenization.

2.2. Training details

We followed the same pretraining process as BART for GreekBART. We utilized text infilling and sentence permutations. Pretraining was conducted on a GPU cluster with a batch size of 768,000 tokens per GPU and an update frequency of 128. The initial learning rate was 6.10^{-4} , and it decreased linearly during the training process. A warm-up phase covering 6% of the total training steps was implemented, while the dropout rate was set to 0.1 for the initial 12 epochs. From epochs 12 to 16, the dropout rate was reduced to 0.05, and finally, it was set to zero from epochs 16 to 20. The experiments were conducted using the Fairseq library (Ott et al., 2019).

3. GreekSUM

Transformer-based Seq2Seq models, including BART, can perform not only extractive but abstractive summarization, as well. This type of summarization is one of the most central and challenging evaluation tasks in NLP. However, there is not any available summarization dataset for the Greek language. Therefore, we created the first dataset in

³<https://dumps.wikimedia.org/elwiki/>

⁴<https://github.com/whitfin/runiq>

the Greek language, well-suited to the abstractive summarization task. Our main goal was to create a Greek version equivalent of the OrangeSum dataset (Kamal Eddine et al., 2021) and XSum dataset (Narayan et al., 2018).

3.1. Data collection & Processing

We scraped data from the "News24/7" website⁵, a prominent news source in Greece, spanning from October 2007 to June 2022. Our dataset included articles in five major categories: politics, society, economy, culture, and world, which were used for two summarization tasks: GreekSUM Title and GreekSUM Abstract.

To ensure data quality, we filtered out empty articles, those with very short titles or abstracts, and eliminated duplicates based on content similarity. We also removed articles with abstracts resembling introductions rather than summaries, discarding 10% of articles with the highest proportion of novel unigrams in the abstracts (threshold of 46.7% novel unigrams). For testing and validation, 10k pairs were reserved for each, with the remaining pairs used for training. The released GreekSUM dataset can be reproduced using our code⁶.

3.2. Analysis

Comparing GreekSUM with OrangeSum, XSum, CNN, DailyMail, and NY Times datasets (Hermann et al., 2015), we observe that GreekSUM and OrangeSum have similar average document and summary lengths. GreekSUM is also comparable in scale to XSum. Extractive methods (*i.e.* LEAD and EXT-ORACLE) show poor performance on GreekSUM, indicating that the dataset is less biased towards extractive models. Thus, GreekSUM appears to be more abstractive than traditional summarization datasets (CNN, DailyMail, NY Times). However, the summaries and titles of GreekSUM exhibit less novelty compared to OrangeSum and XSum, with 20.6% novel unigrams in abstracts and 26.7% in titles (compared to 30% and 26.5% in OrangeSum and 35.7% in XSum). Therefore, the summaries in GreekSUM are not as abstractive as desired.

4. Experiments

In this section, we present the results of our experiments in two types of tasks: discriminative and summarization. For discriminative tasks, we compare GreekBART with BART-random, Greek-BERT (Koutsikakis et al., 2020), and XLM-R (Conneau

⁵<https://www.news247.gr/>

⁶<https://github.com/iakovosevdaimon/GreekSUM>

et al., 2020) models. We fine-tuned all models on the tasks and evaluated them against monolingual and multilingual baselines.

In the summarization task, we compare GreekBART, BART-random, and two versions of mBART (Liu et al., 2020; Tang et al., 2020): mBART25 and mBART50. mBART models are pretrained on multiple languages excluding Greek. Therefore, we performed zero-shot learning for the summarization task. On the other hand, the BART-random model uses the same architecture and vocabulary as GreekBART, however, it is trained from scratch on the downstream tasks.

4.1. Discriminative tasks

Apart from generative tasks, the BART model also demonstrates remarkable performance in discriminative tasks (Lewis et al., 2020). We examine the performance of the models (*i.e.* Greek-BERT, XLM-R, BART-random, GreekBART) on four discriminative tasks:

- **2 NCC tasks (News Category Classification tasks):** For the first classification task, we used the novel summarization dataset, GreekSUM (see section 3). We considered the five distinct subjects that an article may fall into as its labels. Whereas, in the second classification task, we used the proposed Greek classification dataset of Lioudakis et al. (2020), which was created from articles from Makedonia newspaper. The dataset contains 8005 articles from 18 different categories.
- **Natural Language Inference:** Cross-lingual Natural Language Inference Corpus (XNLI) (Conneau et al., 2018) contains pairs of sentences. The objective of this task is to determine whether the first sentence, also known as the premise, entails, contradicts, or is neutral in relation to the second sentence, referred to as the hypothesis.
- **Sentiment Analysis task:** We used a publicly available sentiment analysis dataset⁷ about movies' reviews in Greek. To distinguish between positive and negative reviews, we established a threshold of 3 out of 5. Ratings above this threshold were categorized as positive reviews, otherwise as negative reviews.

4.1.1. Results

Table 3 reports the test set accuracy on the four different tasks. We compare our model with GreekBART (Koutsikakis et al., 2020), XLM-R (Conneau

⁷<https://www.kaggle.com/datasets/nikosfragkis/greek-movies-dataset>

Dataset	% of novel n-grams in gold summary				LEAD			EXT-ORACLE		
	unigrams	bigrams	trigrams	4-grams	R-1	R-2	R-L	R-1	R-2	R-L
CNN	16.75	54.33	72.42	80.37	29.15	11.13	25.95	50.38	28.55	46.58
DailyMail	17.03	53.78	72.14	80.28	40.68	18.36	37.25	55.12	30.55	51.24
NY Times	22.64	55.59	71.93	80.16	31.85	15.86	23.75	52.08	31.59	46.72
XSum	35.76	83.45	95.50	98.49	16.30	01.61	11.95	29.79	08.81	22.65
OrangeSum Title	26.54	66.70	84.18	91.12	19.84	08.11	16.13	31.62	17.06	28.26
OrangeSum Abstract	30.03	67.15	81.94	88.3	22.21	07.00	15.48	38.36	20.87	31.08
GreekSUM Title	26.7	67.9	84.5	91.4	14.68	04.46	14.37	23.36	07.39	23.12
GreekSUM Abstract	20.6	50.8	65.3	73.	17.11	06.17	16.69	34.18	14.17	33.93

Table 2: Degree of abstractivity of GreekSUM compared with that of other datasets. It depicts that GreekSUM follows XSum, and OrangeSum, being more abstractive than traditional summarization datasets.

et al., 2020), and BART-random. For all models, their corresponding *BASE* architecture is used. Among the models, we observe that GreekBART is the best in almost all discriminative tasks, except for the sentiment analysis task, where Greek-BERT achieved the best performance. Generally, it is common for BERT models to perform better than BART models in that kind of tasks. The performance of our model (*i.e.* GreekBART) verifies the results of BART paper (Lewis et al., 2020) that models based on that architecture perform well on both generative and discriminative tasks.

4.2. Summarization

We evaluate our model in two summarization tasks: predicting article titles (GreekSUM Title) and abstracts (GreekSUM Abstract) based on content. We compare GreekBART with mBART25, mBART50, and BART-random, alongside extractive baselines EXT-ORACLE and LEAD (Narayan et al., 2018). The LEAD technique generates a summary by extracting the first N sentences from the document, with N set to 1 in our case. On the other hand, EXT-ORACLE selects the set of sentences from the document that maximizes a specific score, with ROUGE-L being the score used in our implementation. In particular, we extracted the one sentence of the document with the highest ROUGE-L score. Table 4 shows ROUGE (Lin, 2004) and BERTScores (Zhang et al., 2019), calculated using M-BERT (Devlin et al., 2019) and Greek-BERT model.

4.2.1. Results & Human Evaluation

In Table 4 we compare the performance of our models fine-tuned on the summarization task. Even though GreekBART is a BART-*BASE* model and it is compared with BART-*LARGE* models, it is able to achieve better performance than all other models in the task of GreekSUM abstract. Only mBART50 achieves a slightly higher BERTScore than GreekBART when evaluated using the M-BERT model. On the other hand, both mBART models surpass our model in the GreekSUM title task. Although,

even in that task the performance of GreekBART is comparable to one of the two mBART models, both in terms of ROUGE and BERTScore. Our evaluation indicates that mBART50 and GreekBART are the most promising models for the two summarization tasks. It is remarkable the fact that both mBART models, which are not pretrained on the Greek language, are capable to achieve a good performance due to the size of GreekSUM dataset, which contains more than 100k training samples. Interestingly, mBART50 outperforms mBART25 in both ROUGE and BERTScore, possibly due to the additional languages supported.

The human evaluation study was conducted using Best-Worst Scaling (Louveire et al., 2015) to validate the results. 11 native Greek speakers from diverse age groups and educational backgrounds participated. Following Narayan et al. (2018) method, 14 documents were randomly selected from the GreekSUM abstract test set, generating all possible pairs of human-authored (Gold), GreekBART, BART-random, mBART25, and mBART50 summaries, resulting in 140 pairs. Participants assessed accuracy, informativeness, and fluency, choosing the best and worst summaries. Each summary pair was rated by three participants, and scores were calculated based on the percentage of times a summary was chosen as the best minus the percentage chosen as the worst. Results in Table 5 showed Gold as the preferred summary, followed by mBART50 and GreekBART. BART-random received a significantly negative score, indicating inferior performance.

5. Conclusion

We developed GreekBART, the first pretrained Seq2Seq model for Greek, and introduced a new summarization dataset. Our model showed to outperform former state-of-the-art models on 3 out of 4 discriminative tasks and to be on par with BART-*LARGE* models on summarization tasks. Moreover, we presented the potential of zero-shot learning by training a multilingual BART model from

Model	NCC		Sentimental Analysis	XNLI
	News24/7 (ours)	Makedonia (Lioudakis et al., 2020)		
Greek-BERT	92.61 \pm 0.19	89.45 \pm 0.84	86.39 \pm 0.06	78.6 \pm 0.62
XLM-R	93.1 \pm 0.51	89.6 \pm 0.29	85.43 \pm 0.05	78.2 \pm 0.59
BART-random	91.33 \pm 0.17	80.17 \pm 0.09	80.87 \pm 0.12	60.1 \pm 0.43
GreekBART (ours)	93.2 \pm 0.29	91.1 \pm 0.43	85.43 \pm 0.19	78.67 \pm 0.25

Table 3: Results on discriminative tasks. We present the mean accuracy as well as the standard deviation.

	GreekSUM Abstract				GreekSUM Title				
	R-1	R-2	R-L	BertScore	R-1	R-2	R-L	BertScore	
LEAD	17.11	06.17	16.69	72.61/63.56	14.68	04.46	14.37	70/57.13	
EXT-ORACLE	34.18	14.17	33.93	73.89/65.43	23.36	07.39	23.12	70.02/57.33	
BASE	BART-random	13.85	04.47	13.65	72.44/63.27	11.55	03.27	11.42	74.47/62.22
	GreekBART (ours)	16.5	06.13	16.21	73.03/ 64.46	15.35	05.02	15.18	75.78/63.98
LARGE	mBART25	15.07	05.8	14.82	72.75/64.08	16.09	05.58	15.93	76.81/65.38
	mBART50	15.53	06.	15.31	73.07 /64.43	16.1	05.59	15.96	76.81/65.38

Table 4: Results on GreekSUM. Except for ROUGE, we provide also the BertScore. The left-hand BERTScore has calculated using the M-BERT model (Devlin et al., 2019), while the right-hand uses the Greek-BERT (Koutsikakis et al., 2020)

	System	Score
	Gold	45.24
BASE	BART-random	-72.62
	GreekBART (ours)	10.71
LARGE	mBART25	-03.57
	mBART50	20.24

Table 5: The results of human evaluation study

scratch on summarization tasks without being pre-trained on Greek. As future work, we can consider the creation of a more abstractive summarization dataset, the exploration of potential correlations between Greek and the 25 additional languages in mBART50, and the improvement of mBART50’s performance on summarization tasks through affordable language-adaptive pretraining for a limited number of epochs. A demo of our work is openly accessible⁸.

6. Acknowledgements

This research was supported by the ANR chair AML/HELAS (ANR-CHIA-0020-01).

This work was granted access to the HPC resources of IDRIS under the allocation 2022-AD011013750 made by GENCI.

We express our gratitude to all the participants who participated in this human evaluation study. Your contributions have helped us improve our understanding of the topic under investigation and have the potential to make a significant impact on future research and practice. We appreciate your willingness to take part in the study.

⁸<http://nlp.polytechnique.fr/greekbart#greek>

7. Ethical Considerations

The GreekSUM dataset was collected using a Python crawler that respected the *robots.txt* of <http://www.news247.gr>. As the dataset is used only for evaluation purposes the content follows the legal instructions listed on the webpage. For the training of GreekBART we used a cluster of 2 NVIDIA V100 GPUs for 20 days. As with the majority of language models that are based on BART architecture the energy resources required for pre-training models currently are very high and need to be tackled soon (Strubell et al., 2019).

8. Limitations

The proposed GreekSUM dataset that we used for the evaluation of our model is limited to news articles from one website only. Thus, the capability of abstractive summarization of GreekBART is assessed on one domain only. This is because there is a lack of non-English benchmarks and tasks. This is also applicable in the discriminative tasks, where the only available ones for Greek are either sentence classification or natural language inference. While other evaluation datasets do not exist for the Greek language (i.e. Word Sense Disambiguation) or are not available to the public (i.e. Named Entity Recognition dataset).

On the other hand, GreekBART is only compared with extractive summarization methods or with large multi-lingual language models for the summarization task. Since it is the first base model for this language and since the base mBART model does not exist publicly, a fair in-depth comparison of GreekBART with other summarization systems could not be conducted.

9. Bibliographical References

References

- Julien Abadji, Pedro Ortiz Suarez, Laurent Romary, and Benoît Sagot. 2022. [Towards a cleaner document-oriented multilingual crawled corpus](#). In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4344–4355, Marseille, France. European Language Resources Association.
- Vicent Ahuir, Lluís-F. Hurtado, José-Ángel González, and Encarna Segarra. 2021. [Nasca and nases: Two monolingual pre-trained models for abstractive summarization in catalan and spanish](#). *Applied Sciences*, 11(21).
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.
- José Cañete, Gabriel Chaperon, Rodrigo Fuentes, Jou-Hui Ho, Hojin Kang, and Jorge Pérez. 2020. Spanish pre-trained bert model and evaluation data. In *PML4DC at ICLR 2020*.
- 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.
- Alexis Conneau and Guillaume Lample. 2019. Cross-lingual language model pretraining. *Advances in neural information processing systems*, 32.
- Alexis Conneau, Ruty Rinott, Guillaume Lample, Adina Williams, Samuel Bowman, Holger Schwenk, and Veselin Stoyanov. 2018. [XNLI: Evaluating cross-lingual sentence representations](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2475–2485, Brussels, Belgium. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). 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 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Moussa Kamal Eddine, Nadi Tomeh, Nizar Habash, Joseph Le Roux, and Michalis Vazirgiannis. 2022. Arabart: a pretrained arabic sequence-to-sequence model for abstractive summarization. *WANLP 2022*, page 31.
- Karl Moritz Hermann, Tomas Kocisky, 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*, volume 28. Curran Associates, Inc.
- Go Inoue, Bashar Alhafni, Nurpeiis Baimukan, Houda Bouamor, and Nizar Habash. 2021. [The interplay of variant, size, and task type in Arabic pre-trained language models](#). In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 92–104, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.
- Moussa Kamal Eddine, Antoine Tixier, and Michalis Vazirgiannis. 2021. [BARThez: a skilled pre-trained French sequence-to-sequence model](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9369–9390, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Philipp Koehn. 2005. [Europarl: A parallel corpus for statistical machine translation](#). In *Proceedings of Machine Translation Summit X: Papers*, pages 79–86, Phuket, Thailand.
- John Koutsikakis, Ilias Chalkidis, Prodromos Malakasiotis, and Ion Androutsopoulos. 2020. [Greek-bert: The greeks visiting sesame street](#). In *11th Hellenic Conference on Artificial Intelligence, SETN 2020*, page 110–117, New York, NY, USA. Association for Computing Machinery.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. [Imagenet classification with deep convolutional neural networks](#). *Commun. ACM*, 60(6):84–90.
- Taku Kudo and John Richardson. 2018. [SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 66–71, Brussels, Belgium. Association for Computational Linguistics.
- 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.
- 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.
- Michalis Lioudakis, Stamatis Outsios, and Michalis Vazirgiannis. 2020. [An ensemble method for producing word representations focusing on the Greek language](#). In *Proceedings of the 3rd Workshop on Technologies for MT of Low Resource Languages*, pages 99–107, Suzhou, China. Association for Computational Linguistics.
- 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.
- Jordan J. Louviere, Terry N. Flynn, and A. A. J. Marley. 2015. *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge University Press.
- Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric de la Clergerie, Djamé Seddah, and Benoît Sagot. 2020. [CamemBERT: a tasty french language model](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics.
- Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018. [Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1797–1807, Brussels, Belgium. Association for Computational Linguistics.
- OpenAI. 2023. [Gpt-4 technical report](#).
- Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. 2019. [fairseq: A fast, extensible toolkit for sequence modeling](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pages 48–53, Minneapolis, Minnesota. Association for Computational Linguistics.
- Stamatis Outsios, Konstantinos Skianis, Polykarpos Meladianos, Christos Xypolopoulos, and Michalis Vazirgiannis. 2018. Word embeddings from large-scale greek web content. *arXiv preprint arXiv:1810.06694*.
- Katerina Papantoniou and Yannis Tzitzikas. 2020. [Nlp for the greek language: A brief survey](#). In *11th Hellenic Conference on Artificial Intelligence*, SETN 2020, page 101–109, New York, NY, USA. Association for Computing Machinery.
- Pranaydeep Singh and Els Lefever. 2022. [When the student becomes the master: Learning better and smaller monolingual models from mBERT](#). In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4434–4441, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.
- Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2019. [Energy and policy considerations for deep learning in NLP](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3645–3650, Florence, Italy. Association for Computational Linguistics.
- Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, Jiatao Gu, and Angela Fan. 2020. [Multilingual translation with extensible multilingual pretraining and fine-tuning](#).
- Nguyen Luong Tran, Duong Minh Le, and Dat Quoc Nguyen. 2021. [Bartpho: Pre-trained sequence-to-sequence models for vietnamese](#). *CoRR*, abs/2109.09701.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2019. [Bertscore: Evaluating text generation with BERT](#). *CoRR*, abs/1904.09675.

A. GreekSUM Abstract

In this appendix section, we present the reference and model summaries of 5 randomly selected documents from the test set of the GreekSUM Abstract.

Document	«Ο κύβος ερρίφθη. Ο Άρμιν Λάσετ θα είναι ο υποψήφιος Καγκελάριος για την Χριστιανική Ένωση», δήλωσε πριν από λίγο ο Αρχηγός της Χριστιανοκοινωνικής Ένωσης (CSU) και Πρωθυπουργός της Βαυαρίας Μάρκος Ζέντερ, αναγνωρίζοντας το αποτέλεσμα της ψηφοφορίας του προεδρείου του Χριστιανοδημοκρατικού Κόμματος (CDU), το οποίο σε ποσοστό 77,5% τάχθηκε υπέρ της υποψηφιότητας του κ. Λάσετ. Πριν από την συνεδρίαση του προεδρείου του CDU, ο κ. Ζέντερ είχε δηλώσει ότι παραχωρεί στο CDU το προβάδισμα στην επιλογή του υποψήφιου Καγκελάριου της Χριστιανικής Ένωσης (CDU/CSU) και σήμερα επανέλαβε ότι δέχεται το αποτέλεσμα «χωρίς μνησικαϊά» και ότι τάσσεται υπέρ της ενότητας της Χριστιανικής Ένωσης.
Gold	Ο Άρμιν Λάσετ θα είναι ο υποψήφιος των CDU και CSU για την καγκελαρία της Γερμανίας στις εκλογές του Σεπτεμβρίου.
ABSTRACT BART-random	Ο Άρμιν Λάσετ θα είναι ο υποψήφιος πρωθυπουργός της Χριστιανικής Ένωσης, μετά από σχετική συνεδρίαση.
mBART25	Ο πρωθυπουργός της Βαυαρίας δέχθηκε το αποτέλεσμα της ψηφοφορίας του προεδρείου του CDU, το οποίο σε ποσοστό 77,5% τάχθηκε υπέρ της υποψηφιότητας του Άρμιν Λάσετ.
mBART50	Σε ποσοστό 77,5% τάχθηκε υπέρ της υποψηφιότητας του Άρμιν Λάσετ στο προεδρείο του CDU, ο Πρωθυπουργός της Βαυαρίας Μάρκος Ζέντερ.
GreekBART	Υπέρ του Άρμιν Λάσετ τάσσεται ο Μάρκος Ζέντερ, αναγνωρίζοντας το αποτέλεσμα της ψηφοφορίας του προεδρείου του CDU.

Table 6: Example 1-GreekSUM Abstract

Document	Κλειστή είναι η λεωφόρος Βασ. Κωνσταντίνου στο ύψος του Παναθηναϊκού Σταδίου, με αποτέλεσμα να έχει δημιουργηθεί κυκλοφοριακό πρόβλημα, καθώς έχει χυθεί μεγάλη ποσότητα λαδιού από φορτηγό, στην συμβολή με την λεωφόρο Βασ. Όλγας και είναι επικίνδυνη η διέλευση των οχημάτων. Η Τροχαία έχει διακόψει την κυκλοφορία στο κατερχόμενο ρεύμα στο ύψος της οδού Ριζάρη και στο ανερχόμενο από την αρχή της Αρδητού και κάνει εκτροπή, αλλά έχει δημιουργηθεί μπουτιλιάρισμα. Στο σημείο που έχουν χυθεί τα λάδια βρίσκονται συνεργεία του Δήμου, που ρίχνουν πριονίδι και άλλα υλικά για να αντιμετωπίσουν την ολοσθηρότητα του οδοστρώματος και να αποκατασταθεί η κυκλοφορία.
Gold	Η Τροχαία έχει διακόψει την κυκλοφορία στο κατερχόμενο ρεύμα στο ύψος της οδού Ριζάρη και στο ανερχόμενο από την αρχή της Αρδητού και κάνει εκτροπή - Μπουτιλιάρισμα στο σημείο.
ABSTRACT BART-random	Η Τροχαία, που έχει δημιουργηθεί στο ύψος του Παναθηναϊκού, έχει διακοπεί την κυκλοφορία των οχημάτων στην λεωφόρο Βασ.Α.
mBART25	Κλειστή είναι η λεωφόρος Βασ. Κωνσταντίνου στο ύψος του Παναθηναϊκού Σταδίου, με αποτέλεσμα να έχει δημιουργηθεί κυκλοφοριακό πρόβλημα.
mBART50	Κυκλοφοριακό πρόβλημα στο ύψος του Παναθηναϊκού Σταδίου, καθώς έχει χυθεί μεγάλη ποσότητα λαδιού από φορτηγό σε λεωφόρο Βασ. Κωνσταντίνου.
GreekBART	Κυκλοφοριακό πρόβλημα έχει δημιουργηθεί στην Λεωφόρο Βασ. Κωνσταντίνου στο ύψος του Παναθηναϊκού Σταδίου, με αποτέλεσμα να έχει δημιουργηθεί μπουτιλιάρισμα.

Table 7: Example 2-GreekSUM Abstract

Document	<p>Η Καγκελάριος Άνγκελα Μέρκελ δεν θα παραστεί στην επίσημη δεξίωση που θα παραθέσει την Παρασκευή ο Ομοσπονδιακός Πρόεδρος Φρανκ-Βάλτερ Στάϊνμάιερ προς τιμήν του Προέδρου της Τουρκίας Ρετζέπ Ταγίπ Ερντογάν, σύμφωνα με κυβερνητικές πηγές τις οποίες επικαλείται το περιοδικό «Der Spiegel». Η δεξίωση αλλά και οι στρατιωτικές τιμές με τις οποίες θα υποδεχθεί τον προσκεκλημένο του ο Γερμανός Πρόεδρος προκαλούν σοβαρές αντιδράσεις στον πολιτικό κόσμο της χώρας. Η Μέρκελ είναι πάντα προσκεκλημένη του Ομοσπονδιακού Προέδρου σε δεξιώσεις ή επίσημα δείπνα που παρατίθενται προς τιμήν υψηλών προσκεκλημένων. Η ίδια ωστόσο συνηθίζει να παρευρίσκεται μόνο σε εξαιρετικές περιπτώσεις. Η τελευταία φορά που παρέστη σε κάτι ανάλογο ήταν το επίσημο δείπνο που είχε παρατεθεί το 2015 προς τιμήν της Βασίλισσας Ελισάβετ, ενώ την προηγούμενη χρονιά είχε παρευρεθεί στο δείπνο με τον Εμίρη του Κατάρ. Αντιθέτως, δεν είχε παρευρεθεί στην δεξίωση προς τιμήν του Κινέζου Προέδρου Σι Τζινπίνγκ το 2017. Η Καγκελάριος όμως δεν θα είναι η μόνη που θα απορρίψει την πρόσκληση του Στάϊνμάιερ. Ο Πρόεδρος των Φιλελευθέρων (FDP) Κρίστιαν Λίντντερ ανακοίνωσε ότι δεν σκοπεύει να παραστεί, καθώς δεν επιθυμεί «να συμμετάσχει στην προπαγάνδα του Ερντογάν». Την ίδια στάση θα τηρήσει και η εκπρόσωπος του κόμματος για την εξωτερική πολιτική, Μπιτζάν Ντζιρ-Σαράι, ενώ σύσσωμη η ηγετική ομάδα των Πρασίνων, οι συμπρόεδροι Ανάλενα Μπέρμποκ και Ρόμπερτ Χάμπεκ και οι επικεφαλής της Κοινοβουλευτικής Ομάδας Κάτριν Γκέριγκ-Έχαρτ και Άντον Χοφράιτερ, δήλωσαν ότι θα απέχουν από την δεξίωση. Το ίδιο ισχύει και για τους επικεφαλής της Εναλλακτικής για την Γερμανία (AfD) Άλεξάντερ Γκάουλαντ και Αλίσ Βαϊντέλ και για την επικεφαλής της Κ. Ο. της Αριστεράς Σεβίμ Νταγκντελέν. Αντιθέτως, την πρόθεσή του να παραστεί στην δεξίωση στο Προεδρικό Ανάκτορο Bellevue εξέφρασε ο πρώην Πρόεδρος των Πρασίνων Τζεμ Έζντεμιρ, διευκρινίζοντας ταυτόχρονα ότι ο Τούρκος Πρόεδρος «δεν είναι κανονικός Πρόεδρος και δεν αξίζει» να παρατεθεί δεξίωση προς τιμήν του. Με την παρουσία του, δήλωσε ο κ. Έζντεμιρ στην «Tagesspiegel», ελπίζει να στείλει ένα μήνυμα τόσο προς την Τουρκία όσο και προς την τουρκογερμανική κοινότητα: «Η αντιπολίτευση στην Γερμανία είναι μέρος της πολιτικής αυτής της χώρας, είμαστε ένα σταθερό και απαραίτητο συστατικό στοιχείο της δημοκρατίας μας. Ο κ. Ερντογάν θα πρέπει να με ανεχθεί».</p>
<p>Gold</p> <p>BART-random</p> <p>mBART25</p> <p>mBART50</p> <p>GreekBART</p>	<p>Η καγκελάριος είναι πάντα προσκεκλημένη του ομοσπονδιακού προέδρου σε δεξιώσεις ή δείπνα προς τιμήν υψηλών προσκεκλημένων, ωστόσο δίνει το παρών μόνο σε εξαιρετικές περιπτώσεις.</p> <p>Δεν θα παραστεί στην επίσημη δεξίωση που θα παραθέσει την Τουρκία προς τιμήν του Ρετζέπ Ταγίπ Ερντογάν ο εκπρόσωπος της Γερμανίας Άνγκελα Μέρκελ.</p> <p>Αντιδράσεις από τον πολιτικό κόσμο της χώρας προκαλούν η δεξίωση που θα παραθέσει ο Φρανκ-Βάλτερ Στάϊνμάιερ προς τιμήν του Προέδρου της Τουρκίας - Δεν θα είναι η μόνη που θα απορρίψει την πρόσκληση του Στάϊνμάιερ.</p> <p>Η Μέρκελ είναι πάντα προσκεκλημένη του Ομοσπονδιακού Προέδρου σε δεξιώσεις ή επίσημα δείπνα που παρατίθενται προς τιμήν υψηλών προσκεκλημένων. Η ίδια ωστόσο συνηθίζει να παρευρίσκεται μόνο σε εξαιρετικές περιπτώσεις.</p> <p>Από τον πολιτικό κόσμο της Γερμανίας. Η Άνγκελα Μέρκελ δεν θα παραστεί στην επίσημη δεξίωση προς τιμήν του Γερμανού Προέδρου Φρανκ-Βάλτερ Στάϊνμάιερ.</p>

Table 8: Example 3-GreekSUM Abstract

	<p>Document</p> <p>Από το 2011 και μετά αρκετοί εκατοντάδες άνθρωποι έχουν πεθάνει στην προσπάθειά τους να βγάλουν την τέλεια selfie. Οι περισσότεροι θάνατοι έχουν λάβει χώρα στην Ινδία. Ακολουθεί η Ρωσία, οι Ηνωμένες Πολιτείες και ύστερα το Πακιστάν με τους νεκρούς συνολικά να φτάνουν τους 259. Βέβαια υπάρχουν κάποια σημεία, τα οποία σύμφωνα με έρευνες, παρουσιάζουν μεγαλύτερη επικινδυνότητα, όπως το νερό και οι ψηλές κυλιόμενες σκάλες. Οι πιο «συνηθισμένες» αιτίες θανάτου από selfie συμπεριλαμβάνουν τον πνιγμό, την πτώση, τη σύγκρουση με κινούμενο όχημα και τις φωτιές. Όσον αφορά τα στατιστικά στοιχεία τα 3/4 των θυμάτων είναι άνδρες και κάτω από την ηλικία των 30. Αν και οι γυναίκες βγάζουν περισσότερες selfie σύμφωνα με τις μελέτες, οι άνδρες είναι πιο επιρρεπείς στον κίνδυνο. Ακόμα, οι τουρίστες είναι αυτοί που πλήττονται πιο συχνά στην προσπάθεια να βγάλουν μια φωτογραφία που θα εντυπωσιάσει τους ακολούθους τους. Οι αρχές ψάχνουν τρόπους προκειμένου να αποτρέψουν τους θανάτους. Για παράδειγμα η ρωσική αστυνομία μοίρασε φυλλάδια, τα οποία εμπεριείχαν προειδοποιήσεις σχετικά με τους κινδύνους που «καραδοκούν» πίσω από μια selfie. Στις Ηνωμένες Πολιτείες, τα εθνικά πάρκα έχουν εκδώσει οδηγούς για το πώς να βγάζεις «ασφαλείς» selfies, ενώ στην Ινδία υπάρχουν επίσημα σχεδιασμένες πινακίδες που προειδοποιούν για υψηλού κινδύνου περιοχές ή αλλιώς “No selfie zones”. Αν και η εμμονή με τις selfie δεν φαίνεται να περνάει οι αρχές κάνουν ότι μπορούν για να περιορίσουν την επικινδυνότητα και τους θανάτους.</p>
<p>ABSTRACT</p>	<p>Gold</p> <p>Οι πιο «συνηθισμένες» αιτίες θανάτου από selfie συμπεριλαμβάνουν πνιγμό, πτώση, και τη σύγκρουση με κινούμενο όχημα - Άνδρες κάτω των 30 τα περισσότερα θύματα.</p> <p>BART-random</p> <p>Οι Ηνωμένες Πολιτείες, Ινδία, Αν. και Πακιστάν και Αν. Ινδία αναζητούν αναζητούν στοιχεία για να βγάλουν την τέλεια selfie τους στην προσπάθειά τους.</p> <p>mBART25</p> <p>Η Ινδία μετράει τους 259 θανάτους από selfie, τα οποία συμπεριλαμβάνουν τον πνιγμό, την πτώση, τη σύγκρουση με κινούμενο όχημα και τις φωτιές. Οι αρχές ψάχνουν τρόπους προκειμένου να αποτρέψουν τους θανάτους.</p> <p>mBART50</p> <p>Στην Ινδία, τα εθνικά πάρκα έχουν εκδώσει οδηγούς για το πώς να βγάζεις «ασφαλείς» selfies, ενώ στην Ινδία υπάρχουν επίσημα σχεδιασμένες πινακίδες που προειδοποιούν για υψηλού κινδύνου περιοχές.</p> <p>GreekBART</p> <p>Πολλοί άνθρωποι έχουν πεθάνει στην προσπάθειά τους να βγάλουν μια selfie, με τις «συνηθισμένες» αιτίες να συμπεριλαμβάνουν τον πνιγμό, την πτώση, τη σύγκρουση με κινούμενο όχημα και τις φωτιές.</p>

Table 9: Example 4-GreekSUM Abstract

Document	<p>Στην απώλεια του Μίκη Θεοδωράκη αναφέρθηκε ο πρωθυπουργός Κυριάκος Μητσοτάκης στην έναρξη της συνεδρίασης του Υπουργικού Συμβουλίου, κηρύσσοντας τριήμερο εθνικό πένθος. Ο πρωθυπουργός ειδικότερα δήλωσε: “Τη σημερινή μας συνεδρίαση σκιάζει δυστυχώς μία πολύ θλιβερή είδηση: Ο Μίκης Θεοδωράκης περνά πια στην αιωνιότητα. Η φωνή του σίγησε και μαζί του σίγησε και ολόκληρος ο Ελληνισμός. Όπως είχε γραφτεί και για τον Παλαμά, «όλοι είχαμε ξεχάσει πως είναι θνητός». Όμως, μας αφήνει παρακαταθήκη τα τραγούδια του, την πολιτική του δράση, αλλά και την εθνική του προσφορά σε κρίσιμες στιγμές. Η Ρωμιοσύνη σήμερα κλαίει. Και γι’ αυτό και με απόφαση της κυβέρνησης από σήμερα κηρύσσεται τριήμερο εθνικό πένθος. Όπως ξέρετε, είχα την τιμή να τον γνωρίζω για πολλά χρόνια και σχετικά πρόσφατα μάλιστα τον είχα επισκεφτεί. Οι συμβουλές του ήταν πάντα πολύτιμες για μένα, κυρίως αυτές που αφορούσαν στην ενότητα του λαού μας και στην υπέρβαση των διαχωριστικών γραμμών. Πιστεύω πως η καλύτερη τιμή προς αυτόν τον παγκόσμιο Έλληνα θα είναι εμείς, με το καθημερινό μας έργο, να κάνουμε πράξη αυτό ακριβώς το μήνυμά του. Ο Μίκης είναι η Ιστορία μας και πρέπει να τη συνεχίσουμε όπως θα ήθελε και εκείνος.” Πέθανε ο Μίκης Θεοδωράκης - Ορφάνεψε η Ρωμιοσύνη Έλενα Ακρίτα - Ο Ήλιος (που κρύφτηκε) και ο Χρόνος (που χάθηκε), Μίκη Μίκης Θεοδωράκης: Τα 5 τραγούδια του σπουδαίου μουσικού που «μιλούν» στην ψυχή της Ελλάδας Ο πολιτικός Μίκης Θεοδωράκης: Πάντα στο πλευρό των απλών ανθρώπων.</p>
<p style="text-align: center;">Gold</p> <p style="writing-mode: vertical-rl; transform: rotate(180deg);">ABSTRACT</p> <p>BART-random</p> <p>mBART25</p> <p>mBART50</p> <p>GreekBART</p>	<p>Η Ρωμιοσύνη σήμερα κλαίει δήλωσε ο πρωθυπουργός στην έναρξη της συνεδρίασης του υπουργικού συμβουλίου αναφερόμενος στο θάνατο του Μίκη Θεοδωράκη.</p> <p>Ο πρωθυπουργός κατά την έναρξη της συνεδρίασης του Υπουργικού Συμβουλίου κηρύσσοντας την απώλεια του Μίκη Θεοδωράκη.</p> <p>Ο πρωθυπουργός Κυριάκος Μητσοτάκης απο το υπουργικό συμβούλιο για τον θάνατο του Μίκη Θεοδωράκη.</p> <p>Τριήμερο εθνικό πένθος κηρύχθηκε στη συνεδρίαση του υπουργικού συμβουλίου, με τον πρωθυπουργό να σημειώνει ότι ο Μίκης Θεοδωράκης περνά πια στην αιωνιότητα.</p> <p>Το δικό του μήνυμα για την απώλεια του Μίκη Θεοδωράκη έστειλε ο πρωθυπουργός Κυριάκος Μητσοτάκης κατά τη συνεδρίαση του Υπουργικού Συμβουλίου.</p>

Table 10: Example 5-GreekSUM Abstract

B. GreekSUM Title

In the second section of the appendices, we present the reference and model titles of 5 randomly selected documents from the test set of the GreekSUM Title.

Document	Ένας 33χρονος έχασε τη ζωή του, ύστερα από σύγκρουση δύο αυτοκινήτων, έξω από τη Θεσσαλονίκη. Όπως έγινε γνωστό, το θανατηφόρο τροχαίο συνέβη στις 2.15 μετά τα μεσάνυχτα σε παράδρομο της Εγνατίας Οδού, στο ύψος του Ωραιοκάστρου. Σύμφωνα με την Αστυνομία, ο 33χρονος, οδηγός του ενός οχήματος, διακομίστηκε στο νοσοκομείο Παπαγεωργίου, όπου όμως λίγη αργότερα υπέκυψε στα τραύματά του, ενώ η οδηγός του άλλου οχήματος υπέστη ελαφρά τραύματα. Οι ακριβείς συνθήκες υπό τις οποίες προκλήθηκε η σύγκρουση ερευνώνται από το αρμόδιο τμήμα τροχαίας.
Gold	Τροχαίο δυστύχημα στη Θεσσαλονίκη με έναν νεκρό
BART-random	Τροχαίο έξω από τη Θεσσαλονίκη - Δύο τραυματίες
mBART25	Θεσσαλονίκη: Νεκρός 33χρονος ύστερα από σύγκρουση δύο αυτοκινήτων
mBART50	Θεσσαλονίκη: Νεκρός 33χρονος ύστερα από σύγκρουση δύο αυτοκινήτων
GreekBART	Τροχαίο στη Θεσσαλονίκη: Νεκρός 33χρονος σε παράδρομο

Table 11: Example 1-GreekSUM Title

Document	Όλες οι χώρες της Ευρωπαϊκής Ένωσης συμφωνούν ότι δεν θα πληρώσουν τη Ρωσία απευθείας σε ρούβλια για τις εισαγωγές ρωσικού φυσικού αερίου, δήλωσαν υψηλόβαθμοι ευρωπαίοι αξιωματούχοι, σημειώνοντας ότι οι επόμενες πληρωμές είναι προγραμματισμένες για τις 20 Μαΐου. «Αυτό που γνωρίζουμε, και υπάρχει συναίνεση επάνου μεταξύ όλων των κρατών μελών, είναι ότι κανείς δεν είναι πρόθυμος να πληρώσει σε ρούβλια», δήλωσε ο ένας αξιωματούχος κατά την διάρκεια ενημέρωσης των δημοσιογράφων και προσθέτοντας ότι η Ευρωπαϊκή Επιτροπή δεν γνωρίζει πόσοι αγοραστές έχουν ανοίξει λογαριασμούς για πληρωμές προμήθειας φυσικού αερίου μέσω της Gazprombank. Στο μεταξύ, ανώτερος αξιωματούχος της Ευρωπαϊκής Ένωσης δήλωσε πως και μόνο το άνοιγμα τραπεζικού λογαριασμού σε ρούβλια στην Gazprombank ενδέχεται να αποτελεί παραβίαση των κυρώσεων που έχει επιβάλει η ΕΕ σε βάρος της Ρωσίας, όμως η ΕΕ δεν έχει ένδειξη πως κάποια εταιρεία φυσικού αερίου της ΕΕ έχει κάνει κάτι τέτοιο. Ο αξιωματούχος δήλωσε πως «εκ πρώτης όψews» το άνοιγμα τραπεζικών λογαριασμών σε ρούβλια από εισαγωγείς φυσικού αερίου φαίνεται ότι παραβιάζει τις κυρώσεις. Ο αξιωματούχος πρόσθεσε πως η Ευρωπαϊκή Επιτροπή δεν έχει κάποια επίσημη ένδειξη ότι εταιρείες της ΕΕ έχουν δημιουργήσει στην Gazprombank λογαριασμούς σε ρούβλια για την πληρωμή του φυσικού αερίου. Επίσης διευκρίνισε πως η Πολωνία και η Βουλγαρία χρησιμοποίησαν τις υφιστάμενες μεθόδους πληρωμής για το ρωσικό αέριο, πριν η Μόσχα αναστείλει χθες, Τετάρτη, τις προμήθειες των χωρών αυτών με αέριο, και πως δεν χρησιμοποίησαν τον μηχανισμό που προτείνει η Μόσχα για να πληρώσουν σε ρούβλια. «Σύμφωνα με τις πληροφορίες μας, αμφότερες οι χώρες επέμειναν στην αρχική μορφή πληρωμής», δήλωσε ο αξιωματούχος σε δημοσιογράφους. Ωστόσο δύο πηγές είπαν σήμερα στο Ρόιτερς ότι λίγες ευρωπαϊκές εταιρείες έχουν αρχίσει να πληρώνουν σε ρούβλια τη Ρωσία για το φυσικό αέριο, αν και μεγάλοι πελάτες της δεν το έχουν κάνει ακόμη. «Μερικές εμπορικές εταιρείες, ίσως περισσότερες από πέντε, έχουν αρχίσει τις πληρωμές», είπε μία πηγή, ζητώντας να μην κατονομαστεί, επειδή δεν είχε εξουσιοδοτηθεί να μιλήσει στα μέσα ενημέρωσης.
Gold	Φυσικό αέριο: Όλες οι χώρες της ΕΕ συμφωνούν ότι δεν θα πληρώσουν τη Ρωσία σε ρούβλια
BART-random	E.E.: «Δεν θα πληρώσουν» οι χώρες της ΕΕ για το φυσικό αέριο σε ρούβλια
mBART25	EE: Οι χώρες δεν πληρώνουν σε ρούβλια τη Ρωσία για το φυσικό αέριο
mBART50	EE: Οι χώρες δεν πληρώνουν σε ρούβλια τη Ρωσία για το φυσικό αέριο
GreekBART	EE: Δεν θα πληρώσουμε τη Ρωσία σε ρούβλια για το φυσικό αέριο

Table 12: Example 2-GreekSUM Title

Document	<p>Στις ημέρες του Πάσχα έχει προσαρμοστεί το πρόγραμμα λειτουργίας λεωφορείων, τρόλεϊ, ηλεκτρικού και μετρό. Ειδικότερα, τα λεωφορεία και τα τρόλεϊ σήμερα, Μεγάλη Παρασκευή, θα κινούνται με πρόγραμμα Σαββάτου. Οι συρμοί στο μετρό θα διέρχονται από τους σταθμούς ανά 7 λεπτά από τις 09.00 έως τις 17.00 και ανά 10 λεπτά τις υπόλοιπες ώρες. Υπενθυμίζεται πως δεν θα ισχύσει η δίωρη παράταση λειτουργίας που εφαρμόζεται τις Παρασκευές. Στον ηλεκτρικό οι συρμοί θα διέρχονται από τους σταθμούς ανά 10,5 λεπτά. Τα λεωφορεία και τα τρόλεϊ θα κινηθούν με πρόγραμμα Κυριακής, ενώ θα αποσυρθούν νωρίτερα, ώστε να βρίσκονται στα αμαξοστάσια στις 23.00. Τα λεωφορεία θα κινηθούν με πρόγραμμα Κυριακής και τα τρόλεϊ με ειδικό πρόγραμμα Κυριακής. Τόσο στα δρομολόγια των λεωφορείων όσο και σ' αυτά των τρόλεϊ θα εφαρμοστεί ειδικό πρόγραμμα Σαββάτου. Ακινήτοποιημένοι θα μείνουν την Τετάρτη 1η Μαΐου οι συρμοί του ηλεκτρικού (πρώην ΗΣΑΠ), τα λεωφορεία, τα τρόλεϊ, αλλά και ο σιδηρόδρομος, λόγω 24ωρης απεργίας των εργαζομένων, που θα συμμετάσχουν στις απεργιακές συγκεντρώσεις για την Πρωτομαγιά. Όπως αναφέρουν σε ανακοίνωσή τους οι εργαζόμενοι στον πρώην ΗΣΑΠ, «είναι μέρα αγώνα, τιμής και μνήμης. Θυμόμαστε και τιμάμε τους πρωτοπόρους αγωνιστές και τα θύματα των εργατικών αγώνων για βελτίωση των συνθηκών δουλειάς για αξιοπρεπείς αμοιβές και την κατοχύρωση των δικαιωμάτων μας. Ανασυγκροτούμαστε, θέτουμε τους στόχους μας και προχωράμε σε νέους αγώνες. Διεκδικούμε και παλεύουμε για την αναπλήρωση απωλειών από τις μνημονιακές πολιτικές λιτότητας, για πραγματικές αυξήσεις στους μισθούς και στις κοινωνικές παροχές». Και προσθέτουν «υπερασπιζόμαστε τον δημόσιο χαρακτήρα των συγκοινωνιών. Διεκδικούμε την υπογραφή νέας Συλλογικής Σύμβασης Εργασίας. Αγωνιζόμαστε για ασφαλείς, φθηνές συγκοινωνίες. Με αγώνες κατακτάμε τα δικαιώματά μας».</p>
<p>Gold BART-random mBART25 mBART50 GreekBART</p>	<p>Πάσχα 2019: Πώς θα κινηθούν λεωφορεία, τρόλεϊ, ηλεκτρικός και μετρό Μέσα Μαζικής Μεταφοράς: Πώς θα κινηθούν σήμερα τα Μέσα Μεταφοράς Μέσα Πάσχα: Πώς θα κινηθούν σήμερα λεωφορεία, τρόλεϊ, ηλεκτρικό και μετρό Μέσα Πάσχα: Πώς θα κινηθούν σήμερα λεωφορεία, τρόλεϊ, ηλεκτρικό και μετρό Πάσχα: Πώς θα κινηθούν σήμερα λεωφορεία, τρόλεϊ, ηλεκτρικού και μετρό</p>

Table 13: Example 3-GreekSUM Title

Document	<p>Συνάντηση με οικονομικούς παράγοντες από το Σίτι του Λονδίνου έχει αυτή την ώρα ο Αλέξης Τσίπρας στο κέντρο της βρετανικής πρωτεύουσας. Τον Έλληνα πρωθυπουργό υποδέχθηκε ο αντιπρόεδρος της Επιτροπής Πολιτικής του Σίτι, Τομ Σλέι (Tom Sleigh). Επισημαίνεται ότι η Επιτροπή υπέχει θέση Διοίκησης του Σίτι του Λονδίνου. Από την αίθουσα της «Παλιάς Βιβλιοθήκης», ο πρωθυπουργός θα απευθυνθεί σε έναν κύκλο περισσότερων από εκατό σημαίνοντων στελεχών της επενδυτικής/χρηματοπιστωτικής κοινότητας του Σίτι και, σύμφωνα με πληροφορίες, στη συνέχεια θα ακολουθήσει συνάντηση σε πιο στενό κύκλο συμμετεχόντων. Στον απόηχο της απόφασης του Eurogroup για την ελάφρυνση του χρέους, οι επαφές του Αλέξη Τσίπρα με σημαντικούς εκπροσώπους της επενδυτικής/χρηματοπιστωτικής της κοινότητας του οικονομικού κέντρου της Ευρώπης, σηματοδοτούν ένα ευκρινές διεθνές μήνυμα για τις προοπτικές της ελληνικής οικονομίας και της «επόμενης μέρας», στην περίοδο μετά την ολοκλήρωση των μνημονίων. Όπως ανέφερε κυβερνητικός αξιωματούχος, οι σημερινές συναντήσεις είναι ένας σημαντικός σταθμός σε μια «αλυσίδα» επαφών και συνομιλιών που θα συνεχιστούν στο αμέσως επόμενο διάστημα των καλοκαιρινών μηνών και το φθινόπωρο. Ενδεικτική της ευνοϊκής συγκυρίας για την ελληνική οικονομία και το στοίχημα της ανάκαμψης, η χθεσινόβραδινή αναβάθμιση, από τον αμερικανικό οίκο αξιολόγησης Standard & Poor's της μακροπρόθεσμης πιστοληπτικής ικανότητας της χώρας σε B+, χαιρετίζοντας την απόφαση του Eurogroup. Στις 18:00 το απόγευμα ώρα Ελλάδας, ο πρωθυπουργός θα περάσει το κατώφλι της Downing Street 10 προκειμένου να συναντηθεί με την πρωθυπουργό της Βρετανίας, Τερέζα Μέι. Στη συνέχεια θα έχει συνάντηση με τον αρχηγό του Εργατικού Κόμματος, Τζέρεμι Κόρμπιν.</p>
Gold BART-random mBART25 mBART50 GreekBART	<p>Συνάντηση με οικονομικούς παράγοντες από το Σίτι του Λονδίνου έχει ο Αλέξης Τσίπρας</p> <p>Μήνυμα Τσίπρα στο Λονδίνο για το χρέος</p> <p>Συνάντηση Τσίπρα με οικονομικούς παράγοντες στο Σίτι</p> <p>Συνάντηση Τσίπρα με οικονομικούς παράγοντες στο Σίτι</p> <p>Βλέμματα στο Λονδίνο για την ελληνική οικονομία</p>

Table 14: Example 4-GreekSUM Title

Document	<p>Επιβατικό τρένο εκτροχιάστηκε σήμερα περίπου 20 χλμ. βόρεια της Ραμπάτ, με αποτέλεσμα να σκοτωθούν έξι άνθρωποι και άλλοι 86 να τραυματιστούν, σύμφωνα με επίσημο απολογισμό που ανακοινώθηκε στον τόπο του δυστυχήματος. «Ο εκτροχιασμός προκάλεσε έξι θανάτους, σύμφωνα με τον τρέχοντα απολογισμό, και 86 τραυματίες σε σοβαρή κατάσταση», δήλωσε ο Μοχάμεντ Ραμί Ραχίλ, γενικός διευθυντής της εταιρίας σιδηροδρόμων ONCF, ο οποίος μετέβη επί τόπου. «Ξεκίνησε έρευνα για τον προσδιορισμό των αιτιών του δυστυχήματος», πρόσθεσε, σε βίντεο που αναρτήθηκε στα μέσα κοινωνικής δικτύωσης. Θεαματικές εικόνες του δυστυχήματος, που σημειώθηκε γύρω στις 13:00 ώρα Ελλάδας, περίπου 20 χλμ. βόρεια της πρωτεύουσας Ραμπάτ, στο ύψος της κοινότητας Σιντί Μπουκναντέλ, κάνουν τον γύρο των μέσων κοινωνικής δικτύωσης, που είναι πολύ επικριτικά εναντίον της ONCF. Οι εικόνες δείχνουν πολλά βαγόνια εκτροχιασμένα κοντά σε μια γέφυρα στους αγρούς, ενώ η μηχανή είναι πλήρως κατεστραμμένη. Ο οδηγός της αμαξοστοιχίας είναι νεκρός, σύμφωνα με πολλά τοπικά ΜΜΕ. Ο βασιλιάς αποφάσισε να αναλάβει τα έξοδα της κηδείας των θυμάτων και οι τραυματίες θα διακομιστούν στο στρατιωτικό νοσοκομείο της Ραμπάτ με βασιλικές οδηγίες, αναφέρεται σε ανακοίνωση του γραφείου του βασιλιά.</p>
Gold BART-random mBART25 mBART50 GreekBART	<p>Εκτροχιασμός τρένου στο Μαρόκο: Στους 6 οι νεκροί - 86 τραυματίες</p> <p>Ραμπάτ: 20 νεκροί από εκτροχιασμό τρένου</p> <p>ΗΠΑ: Επιβατικό τρένο εκτροχιάστηκε - Έξι νεκροί και 86 τραυματίες</p> <p>ΗΠΑ: Επιβατικό τρένο εκτροχιάστηκε - Έξι νεκροί και 86 τραυματίες</p> <p>Εκτροχιασμός τρένου στη Ραμπάτ: Έξι νεκροί και 86 τραυματίες</p>

Table 15: Example 5-GreekSUM Title