

BARThez: a Skilled Pretrained French Sequence-to-Sequence Model

Moussa Kamal Eddine
École Polytechnique

Antoine J.-P. Tixier
École Polytechnique

Michalis Vazirgiannis
École Polytechnique

Abstract

Inductive transfer learning has taken the entire NLP field by storm, with models such as BERT and BART setting new state of the art on countless NLU tasks. However, most of the available models and research have been conducted for English. In this work, we introduce BARThez, the first large-scale pretrained seq2seq model for French. Being based on BART, BARThez is particularly well-suited for generative tasks. We evaluate BARThez on five discriminative tasks from the FLUE benchmark and two generative tasks from a novel summarization dataset, OrangeSum, that we created for this research. We show BARThez to be very competitive with state-of-the-art BERT-based French language models such as CamemBERT and FlauBERT. We also continue the pretraining of a multilingual BART on BARThez’ corpus, and show our resulting model, mBARThez, to significantly boost BARThez’ generative performance. Code, data and models are publicly available.

1 Introduction

Inductive transfer learning, that is, solving tasks with models that have been pretrained on very large amounts of data, was a game changer in computer vision (Krizhevsky et al., 2012). In NLP, while annotated data are scarce, raw text is virtually unlimited and readily available. It thus emerged that the ability to learn good representations from plain text could greatly improve general natural language understanding.

Trained on gigantic amounts of raw data and with hundreds of GPUs, models based on the Transformer architecture (Vaswani et al., 2017), such as GPT (rad) and BERT (Devlin et al., 2019), have set new state-of-the-art performance in every NLU task. Moreover, users around the world can easily benefit from these improvements, by finetuning the publicly available pretrained models to their

specific applications. This also saves considerable amounts of time, resources and energy, compared with training models from scratch.

BART (Lewis et al., 2020) combined a BERT-like bidirectional encoder with a GPT-like forward decoder, and pretrained this seq2seq architecture as a denoising autoencoder with a more general formulation of the masked language modeling objectives of BERT. Since not only BART’s encoder but also its decoder is pretrained, BART excels on tasks involving text generation.

While the aforementioned efforts have made great strides, most of the research and resources were dedicated to the English language, despite a few notable exceptions. In this paper, we partly address this limitation by contributing BARThez¹, the first pretrained seq2seq model for French.

BARThez, based on BART, was pretrained on a very large monolingual French corpus from past research that we adapted to suit BART’s specific perturbation schemes. Unlike already existing BERT-based French language models such as CamemBERT (Martin et al., 2020) and FlauBERT (Le et al., 2020), BARThez is particularly well-suited for generative tasks. We evaluate BARThez on five sentiment analysis, paraphrase identification, and natural language inference tasks from the recent FLUE benchmark, and two generative tasks from a novel French summarization dataset, OrangeSum, that we created for this research. We also continue the pretraining of an already pretrained multilingual BART on BARThez’ corpus. Our resulting model, mBARThez, significantly boosts BARThez’ performance on generative tasks.

Our contributions are as follows:

- We publicly release the first large-scale pretrained seq2seq model dedicated to the French language, BARThez, featuring 165M parameters, and trained on 101 GB of text for 60 hours with 128 GPUs.

¹named after a legendary French goalkeeper, Fabien Barthez: https://en.wikipedia.org/wiki/Fabien_Barthez

We evaluate BARThez on five discriminative tasks and two generative tasks, with automated and human evaluation, and show that BARThez is very competitive with the state of the art.

- To address the lack of generative tasks in the existing FLUE benchmark, we put together a novel dataset for summarization in French, OrangeSum, that we publicly release² and analyze in this paper. OrangeSum is more abstractive than traditional summarization datasets, and can be considered the French equivalent of XSum (Narayan et al., 2018).
- We continue the pretraining of a multilingual BART on BARThez’ corpus, and show that our resulting model, named mBARThez, offers a significant boost over BARThez on generative tasks.
- We publicly release our code and models³. Our models were also integrated into the highly-popular Hugging Face `Transformers` library⁴. As such, they can easily be distributed and deployed for research or production within a standard, industrial-strength framework. They also have their own APIs and can be interactively tested online.

2 Related work

Learning without labels is enabled via self-supervised learning⁵, a setting in which a system learns to predict part of its input from other parts of its input. In practice, one or more supervised tasks are created from the unlabeled data, and the model learns to solve these tasks with custom objectives.

Some of the earliest and most famous self-supervised representation learning approaches in NLP are word2vec (Mikolov et al., 2013), GloVe (Pennington et al., 2014) and FastText (Bojanowski et al., 2017). While these methods were significant advancements, they produce static representations, which is a major limitation, as words have different meanings depending on the unique contexts in which they are used.

Deep pretrained language models. ELMo (Peters et al., 2018) provided the first contextualized embeddings, by extracting and combining the internal states of a pretrained deep bi-LSTM language model. Except for the word embeddings and the softmax layer, the forwards and backwards RNNs have different parameters. The authors of ELMo showed that the learned representations could be

transferred with great benefits to downstream architectures, to solve a variety of supervised NLU tasks.

Beyond simply combining internal states, Howard and Ruder (2018) proposed ULMFiT, a universal transfer learning method for text classification where the language model is pretrained on a large, general dataset, finetuned on a specific dataset, and finally augmented with classification layers trained from scratch on downstream tasks.

With the OpenAI GPT, rad capitalized on the Transformer architecture (Vaswani et al., 2017), superior and conceptually simpler than recurrent neural networks. More precisely, they pretrained a left-to-right Transformer decoder as a general language model, and finetuned it on 12 language understanding tasks by applying different transformations to the input.

By combining ideas from all the aforementioned models, and introducing bidirectional pretraining, BERT (Devlin et al., 2019) disrupted the NLP field by setting new state-of-the-art performance on 11 NLU tasks, with very wide margins. More precisely, BERT uses a bidirectional Transformer encoder with a masked language model objective, making the learned representations capture both the left and the right contexts, instead of just the left context. The sheer size of BERT, with up to 24 Transformer blocks, plays a role in performance too.

With GPT-2, a version of GPT with over an order of magnitude more parameters than GPT, Radford et al. (2019) showed that as long as they have very large capacities, general language models can reach reasonable performance on many specific NLU tasks out-of-the-box, without any finetuning, i.e., accomplish zero-shot transfer. This demonstrates the fundamental nature and importance of the language modeling objective for inductive transfer learning.

In RoBERTa, Liu et al. (2019) showed that the performance of BERT could be improved by optimizing its hyperparameters and training procedure. The study of why and how BERT works so well has now its own dedicated research field, known as BERTology (Rogers et al., 2020).

Languages. Following the success of BERT for the English language, some BERT models were pretrained and evaluated in other languages. Some examples include Arabic (Antoun et al., 2020), Dutch (de Vries et al., 2019; Delobelle et al., 2020),

²<https://github.com/Tixierae/OrangeSum>

³<https://github.com/moussaKam/BARThez>

⁴<https://huggingface.co/moussaKam>

⁵a term coined by Yann LeCun.

French (Martin et al., 2020; Le et al., 2020), Italian (Polignano et al., 2019), Portuguese (Souza et al., 2019), Russian (Kuratov and Arkhipov, 2019), and Spanish (Cañete et al., 2020).

In addition to the aforelisted monolingual models, multilingual models were also proposed, notably mBERT (Devlin et al., 2019), XLM (Conneau and Lample, 2019) and XLM-R (Conneau et al., 2020).

Abstractive summarization. Abstractive summarization is an important and challenging task, requiring diverse and complex natural language understanding and generation capabilities. A good summarization model needs to read, comprehend, and write well.

GPT-2 can be used for summarization, by sampling a certain numbers of tokens from a given start seed. However, while the generated text is grammatical and fluent, summarization performance is only slightly superior to that of a random extractive baseline.

Being a bidirectional encoder, BERT cannot be used out-of-the-box for language generation, unlike GPT-2. Furthermore, BERT produces single-sentence representations, whereas for summarization, reasoning over multiple sentence and paragraph representations is necessary. Liu and Lapata (2019) proposed a way to overcome these challenges. At the input level, they introduced special tokens to encode individual sentences, interval segment embeddings, and used more position embeddings than in BERT. Then, they combined a pretrained BERT encoder with a Transformer-based decoder initialized at random and jointly trained the two models with different optimizers and learning rates.

BART and mBART. BART (Lewis et al., 2020) is a denoising auto-encoder that jointly pretrains a bidirectional encoder (like in BERT) and a forward decoder (like in GPT) by learning to reconstruct a corrupted input sequence. Both the encoder and the decoder are Transformers. Since not only the encoder but also the decoder is pretrained, BART is particularly effective when applied to text generation tasks.

Liu et al. (2020) pretrained a multilingual BART (mBART) on 25 different languages. They showed that this multilingual pretraining brings significant performance gains on a variety of machine translation tasks. MASS (Song et al., 2019) is another multilingual pretrained sequence to sequence

model, that learns to predict a masked span in the input sequence. The main difference between MASS and BART, is that the former only predicts the masked fragment of the sentence, while the latter learns to reconstruct the entire corrupted sentence. This difference makes MASS less effective in discriminative tasks, given that only the masked span is fed to the decoder (Lewis et al., 2020). ProphetNet (Qi et al., 2020) which also adopts the encoder-decoder structure, introduces a new learning objective called future n-gram prediction. This objective reduces overfitting on local correlations by learning to predict the next n-grams (instead of unigrams) at each time step given the previous context

3 BARThez

Our model is based on BART (Lewis et al., 2020), a denoising auto-encoder. It consists of a bidirectional encoder and a left-to-right auto-regressive decoder.

3.1 Architecture

We use the BASE architecture, with 6 encoder and 6 decoder layers. We did not opt for a LARGE architecture due to resource limitations. Our BASE architecture uses 768 hidden dimensions and 12 attention heads in both the encoder and the decoder. In total, our model has roughly 165M parameters. The architecture has two differences compared with the vanilla seq2seq Transformer (Vaswani et al., 2017). The first one is the use of GeLUs activation layers instead of ReLUs, and the second is the presence of a normalization layer on top of the encoder and the decoder, following Liu et al. (2020). These additional layers help stabilizing the training when using FP16 precision.

3.2 Vocabulary

To generate our vocabulary, we use SentencePiece (Kudo and Richardson, 2018) that implements byte-pair-encoding (BPE) (Sennrich et al., 2016). We do not perform any type of pre-tokenization and we fix the size of the vocabulary to 50K sub-words. The SentencePiece model is trained on a 10GB random sample of the pretraining corpus. We fix the character coverage to 99.95%.

3.3 Self-supervised learning

We use the same pretraining as in BART. That is, BARThez learns to reconstruct an input sequence that was perturbed with a noise function. BARThez'

objective is to minimize the cross-entropy between the predicted and the original text. Formally, having a set of documents $\{X_1, X_2, \dots, X_n\}$ and a noising function n , we aim at finding the parameters θ that minimize:

$$L_\theta = - \sum_i \log P(X_i | n(X_i); \theta)$$

Following [Lewis et al. \(2020\)](#), we apply two types of noise in n . First, we use *text infilling*, where a number of text spans are sampled and replaced with one [MASK] special token. The length of the spans is sampled from a Poisson distribution with ($\lambda = 3.5$) and 30% of the text is masked. Second, we use *sentence permutation*, where the input document, seen as a list of sentences, is shuffled.

3.4 Pretraining corpus

We created a version of FlauBERT’s corpus ([Le et al., 2020](#)) suitable for the two perturbation schemes described in subsection 3.3. Indeed, in the original FlauBERT corpus, each sentence is seen as an independent instance, while in our case, we need complete documents as instances.

Other than that, BARThez’ corpus is similar to FlauBERT’s. It primarily consists in the French part of CommonCrawl, NewsCrawl, Wikipedia and other smaller corpora that are listed in Table 1. To clean the corpus from noisy examples, we used the script⁶ provided by [Le et al. \(2020\)](#). Note that we disabled the Moses tokenizer, as we used SentencePiece, which does not require any pre-tokenization. The total corpus size was 66/101 GB before/after SentencePiece tokenization.

Corpus	Size
CommonCrawl	42.0
NewsCrawl (Li et al., 2019)	9.6
Wikipedia	4.0
GIGA (Tiedemann, 2012)	3.8
ORTOLANG (ATILF and CLLE, 2020)	2.7
MultiUn (Eisele and Chen, 2010)	2.2
EU Bookshop (Skadınš et al., 2014)	2.1

Table 1: BARThez’ corpora (in GB, after cleaning).

3.5 Training details

We pretrained BARThez on 128 NVidia V100 GPUs. We fixed the batch size to 6000 tokens per GPU and the update frequency to 2, which gave a total number of roughly 22k documents per update. We used the Adam optimizer ([Kingma](#)

⁶<https://github.com/getalp/Flaubert>

and [Ba, 2015](#)) with $\epsilon = 10^{-6}$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$, with a learning rate starting from 6.10^{-4} and decreasing linearly as a function of the training step. We used a warm up of 6% of the total number of training steps. Pretraining lasted for approximately 60 hours, allowing for 20 passes over the whole corpus. In the first 12 epochs, we fixed the dropout to 0.1, for epochs 12 to 16 we decreased it to 0.05, and finally we set it to zero for epochs 16 to 20. All experiments were carried out using the Fairseq library ([Ott et al., 2019](#)).

4 mBARThez

mBART ([Liu et al., 2020](#)) is a multilingual BART. It follows a LARGE architecture, with 12 layers in both the encoder and the decoder, hidden vectors of size 1024, and 16 attention heads. It was trained on a multilingual corpus containing 1369 GB of raw text, for over 2.5 weeks on 256 Nvidia V100 GPUs. The multilingual corpus covers 25 different languages, including 56 GB of French text. In the original paper, the authors evaluated mBART on machine translation. However, mBART can also be used to perform monolingual tasks.

We continued the pretraining of the pretrained mBART on BARThez’ corpus (see subsection 3.4) for about 30 hours on 128 Nvidia V100 GPUs, which allowed for 4 passes over BARThez’ corpus. This can be seen as an instance of *language-adaptive pretraining*, which goes a step further than *domain-adaptive pretraining* ([Gururangan et al., 2020](#)). The initial learning rate was set to 10^{-4} and linearly decreased towards zero. We call the resulting model mBARThez.

Note that we eliminated the tokens containing non-latin characters from mBARThez’ embedding layers, reducing its number of parameters from 610M to 458M.

5 OrangeSum

BART-based models are particularly well-suited to generative tasks, but FLUE ([Le et al., 2020](#)), the French equivalent of GLUE, only contains discriminative tasks⁷ ([Wang et al., 2019b](#)).

We therefore decided to create one such task. We opted for single-document abstractive summarization, as it is a generative task that also requires the model to encode its input very well. In other words, for a model to summarize well, it needs

⁷There is no generative task in GLUE or superGLUE ([Wang et al., 2019a](#)) either.

Dataset	Train/val/test	Avg. doc length		Avg. summary length		Vocab size	
		Words	Sentences	Words	Sentences	Docs	Summaries
CNN	90.3/1.22/1.09	760.50	33.98	45.70	3.58	34	89
DailyMail	197/12.15/10.40	653.33	29.33	54.65	3.86	564	180
NY Times	590/32.73/32.73	800.04	35.55	45.54	2.44	1233	293
XSum	204/11.33/11.33	431.07	19.77	23.26	1.00	399	81
OrangeSum Title	30.6/1.5/1.5	315.31	10.87	11.42	1.00	483	43
OrangeSum Abstract	21.4/1.5/1.5	350	12.06	32.12	1.43	420	71

Table 2: Sizes (column 2) are given in thousands of documents. Document and summary lengths are in words. Vocab sizes are in thousands of tokens.

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	1.61	11.95	29.79	8.81	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

Table 3: Degree of abstractivity of OrangeSum compared with that of other datasets, as reported in Narayan et al. (2018). It can be observed that XSum and OrangeSum are more abstractive than traditional summarization datasets.

to both read, comprehend, and write well, making abstractive summarization one of the most central and challenging evaluation tasks in NLP.

Motivation. Our strategy here was to create a French equivalent of the recently introduced XSum dataset (Narayan et al., 2018). XSum was created by scraping articles and their one-sentence summaries from the BBC website, where the summaries are not catchy headlines, but capture the gist of the articles. Unlike the historical CNN, DailyMail, and NY Times datasets (Hermann et al., 2015), which favor extractivity, performing well on XSum requires a high degree of abstractivity.

Data collection. We adopted an analogous strategy, and scraped the “Orange Actu” website⁸. Orange S.A. is a large French telecommunications corporation, with 266M customers worldwide. Our scraped pages cover almost a decade from Feb 2011 to Sep 2020. They belong to five main categories: France, world, politics, automotive, and society⁹. The society category is divided into 8 subcategories: health, environment, people, culture, media, high-tech, unusual (“insolite” in French), and miscellaneous.

Each article featured a single-sentence title as well as a very brief abstract, both professionally written by the author of the article. We extracted these two fields from each page, thus creating two summarization tasks: OrangeSum Title and Or-

angeSum Abstract. Gold summaries are respectively 11.42 and 32.12 words in length on average (see Table 2). Note that like in XSum, titles in OrangeSum tend not to be catchy headlines but rather convey the essence of the article. The same can be said about the abstracts.

Post-processing. As a post-processing step, we removed all empty articles, and articles whose titles were shorter than 5 words. For OrangeSum Abstract, we removed the top 10% articles in terms of proportion of novel unigrams in the abstracts, as we observed that such abstracts tended to be introductions rather than real abstracts. This corresponded to a threshold of 57% novel unigrams.

For both OrangeSum Title and OrangeSum Abstract, we set aside 1500 pairs for testing, 1500 for validation, and used all the remaining ones for training. We make the dataset publicly available¹⁰. An example document with its summaries is provided in Table 4. More examples are available in appendix.

Analysis. Table 2 compares OrangeSum with XSum and the well-known CNN, DailyMail, and NY Times datasets. We can see that the two OrangeSum datasets are very similar to XSum in terms of statistics, but is one order of magnitude smaller than XSum. However, the size of OrangeSum still allows for effective finetuning, as we later demonstrate in our experiments.

Table 3 provides empirical evidence showing that like XSum, OrangeSum is less biased towards

⁸<https://actu.orange.fr/>, ‘Actu’ means News.

⁹root URLs are <https://actu.orange.fr/> for all categories except <https://auto.orange.fr/news/> for automotive.

¹⁰<https://github.com/Tixierae/OrangeSum>

Document	Le 18 octobre dernier, Jacline Mouraud se faisait connaître en publiant sur Facebook une vidéo dans laquelle elle poussait un “coup de gueule” contre le gouvernement. Aujourd’hui, la Bretonne a pris ses distances par rapport au mouvement, notamment face à d’autres figures plus radicales comme Éric Drouet. Jacline Mouraud réfléchit désormais à créer son propre parti, “la seule chose envisageable”, comme elle l’explique au JDD. Nicolas Sarkozy, “le seul qui a des couilles”. Cette figure des “gilets jaunes”, accusée de faire le jeu de LREM estime que “le problème” d’Emmanuel Macron “c’est qu’il est jeune”. “Il devrait y avoir un âge minimum pour être président : 50 ans”, souligne Jacline Mouraud. Dans le JDD, elle raconte d’ailleurs avoir voté blanc lors de la dernière présidentielle. En 2007 et 2012, c’est Nicolas Sarkozy, “le seul qui a des couilles”, que la figure des “gilets jaunes” avait soutenu. En attendant de se lancer, pas question pour elle en tous les cas d’être candidate aux européennes sur une liste de La République en marche.
ABSTRACT	<p>Gold L’une des figures du mouvement ne sera toutefois pas candidate aux prochaines élections européennes.</p> <p>mBART Jacline Mouraud, figure des “gilets jaunes”, estime que le président d’Emmanuel Macron est trop jeune pour être président.</p> <p>mBARThez Dans un entretien au JDD, la figure des “gilets jaunes” Jacline Mouraud révèle qu’elle réfléchit à créer son propre parti.</p> <p>BARThez Dans les colonnes du JDD, la figure des “gilets jaunes” explique qu’elle envisage de se présenter aux européennes sur une liste La République en marche.</p> <p>C2C Retirée de la vie politique depuis plusieurs mois, Bretone Mouraud envisage de se lancer en politique. Et elle réfléchit à quelque chose de plus, rapporte le JDD.</p>
TITLE	<p>Gold “Gilets jaunes” : Jacline Mouraud réfléchit à créer son parti</p> <p>mBART “Gilets jaunes” : Jacline Mouraud lance son propre parti</p> <p>mBARThez “Gilets jaunes” : Jacline Mouraud prend ses distances</p> <p>BARThez La figure des “gilets jaunes” Jacline Mouraud va créer son propre parti</p> <p>C2C “Gilets jaunes” : Jacline Mouraud réfléchit à sa propre candidature</p>

Table 4: Doc 19233 from OrangeSum’s test set, and associated summaries. Incorrect information in orange. C2C stands for CamemBERT2CamemBERT.

extractive systems compared with the traditional datasets used for abstractive summarization. There are 30% novel unigrams in the OrangeSum Abstract reference summaries and 26.5% in OrangeSum Title, compared with 35.7% in Xsum, 17% in CNN, 17% in DailyMail, and 23% in NY Times. This indicates that XSum and OrangeSum summaries are more abstractive. These observations are also confirmed by the fact that the two extractive baselines LEAD and EXT-ORACLE perform much more poorly on XSum and OrangeSum than on the other datasets.

6 Experiments

We compare BARThez and mBARThez with the following models, summarized in Table 5.

- **mBART**. The multilingual BART LARGE described in section 4.
- **CamemBERT2CamemBERT (C2C)**. Following BERT2BERT (Rothe et al., 2020), we finetuned a sequence-to-sequence model whose both encoder and decoder parameters were initialized with CamemBERT LARGE weights. The only weights that were initialized randomly were the encoder-decoder attention weights.
- **BART-random**. Same architecture and vocabulary as BARThez, trained from scratch on the

downstream tasks.

6.1 Summarization

All pretrained models were finetuned for 30 epochs and we used a learning rate that warmed up to 0.0001 (6% of the training steps) and then decreased linearly to 0. BART-random was trained for 60 epochs. We selected the checkpoint associated with the best validation score to generate the test set summaries, using beam-search with a beam size of 4.

We report ROUGE-1-2-L scores (Lin, 2004) and BERTScore scores (Zhang et al., 2020) in Table 6.

Following Narayan et al. (2018), we included two extractive baselines, LEAD and EXT-ORACLE, which respectively extract the first sentence and the set of sentences maximizing ROUGE-L from the document.

Quantitative results. Table 6 compares the performance of the models finetuned on the summarization task. While having four times less parameters, BARThez is on par with mBART, both in terms of ROUGE and BERTScore.

mBARThez provides a significant boost over BARThez and mBART and reaches best performance everywhere. This highlights the importance of adapting a multilingual pretrained model

		Layers	Params	Vocab. size	Pretraining hours	Pretraining GPUs	Corpus size
BASE	BART-random	12	165	50	0	NA	NA
	BARThez (ours)	12	165	50	60	128	66
LARGE	C2C	24	274	32	24	256	138
	mBART	24	610	250	432	256	1369
	mBARThez (ours)	24	458	101	30	256 + 128	1369 + 66

Table 5: Summary of the models used in our experiments. Parameters are given in millions, vocab sizes in thousands, and corpus sizes in GB. C2C stands for CamemBERT2CamemBERT.

	Abstract				Title				
	R-1	R-2	R-L	BertScore	R-1	R-2	R-L	BertScore	
LEAD	22.21	07.00	15.48	14.66/68.02	19.84	08.11	16.13	15.75/68.43	
EXT-ORACLE	38.36	20.87	31.08	28.99/73.39	31.62	17.06	28.26	25.15/71.95	
BASE	BART-random	27.67	08.23	18.50	22.53/70.97	28.76	13.15	25.20	29.67/73.65
	BARThez (ours)	31.44	12.77	22.23	27.51/72.84	40.86	23.68	36.03	40.61/77.74
LARGE	CamemBERT2CamemBERT	29.23	09.79	19.95	25.53/72.09	34.92	18.04	30.83	36.40/76.17
	mBART	31.85	13.10	22.35	27.80/72.94	40.74	23.70	36.04	40.42/77.67
	mBARThez (ours)	32.67	13.73	23.18	28.80/73.32	41.08	24.11	36.41	41.42/78.05

Table 6: Results on OrangeSum. The two BertScore scores are with/without rescaling (Zhang et al., 2020).

to a specific language before finetuning (*language-adaptive pretraining*). This also suggests that, when proper adaptation is conducted, it can be advantageous to capitalize on a multilingual model to perform monolingual downstream tasks, probably because there are some translingual features and patterns to be learned.

Finally, all BART-based models outperform CamemBERT2CamemBERT by a significant margin.

Human evaluation. To validate our positive quantitative results, we conducted a human evaluation study with 11 French native speakers. The evaluators were PhD students from the computer science department of our university, working in NLP and other fields of AI. They volunteered after receiving an email announcement. Following Narayan et al. (2018), we used *Best-Worst Scaling* (Louviere et al., 2015). Two summaries from two different systems, along with their input document, were presented to a human annotator who had to decide which one was *better*. We asked evaluators to base their judgments on *accuracy* (does the summary contain accurate facts?), *informativeness* (is important information captured?) and *fluency* (is the summary written in well-formed French?).

We randomly sampled 14 documents from the test set of OrangeSum Abstract, and for each document, generated all possible pairs of BARThez, mBARThez, mBART, C2C, and GT summaries, resulting in 140 pairs. Each pair was randomly assigned to three different annotators, resulting in 420 evaluation tasks in total. The final score of a model was given as the percentage of time its summary was chosen as *best* minus the percentage of time

it was chosen as *worst*. Scores are reported in Table 10. mBARThez reaches first place, like for the quantitative results, but with an even wider margin. It is also interesting to note that BARThez, which was on par with mBART quantitatively, significantly outperforms it terms of human evaluations. The negative score of CamemBERT2CamemBERT means that its summaries were judged to be worse more often than not.

Surprisingly, BARThez and mBARThez’ summaries were often judged better than the ground truth ones. This could be explained by the fact that the GT summaries, while well-written, may contain information that is missing from the documents, such as dates. Annotators may consider such information as inaccurate (*model hallucinations*).

Qualitative results. As shown in Table 7, mBARThez is more abstractive than BARThez and mBART. E.g., mBARThez introduces on average 15.48% of novel unigrams in its summaries for the Abstract task, compared with 10.93 and 13.40 for BARThez and mBART, respectively. It is interesting to note that despite this superior abstractivity, mBARThez still reaches first place everywhere in terms of the ROUGE metric, which measures n-gram overlap. We hypothesize that BARThez is less abstractive than mBART and mBARThez because it has four times less parameters (BASE vs LARGE architecture).

Finally, it is also to be noted that CamemBERT2CamemBERT (C2C) introduces many new words, which could be considered a good thing at first. However, it also repeats itself a lot (see Table 9) and has low ROUGE, BERTSum, and human

		OrangeSum Abstract				OrangeSum Title			
		1-grams	2-grams	3-grams	4-grams	1-grams	2-grams	3-grams	4-grams
Gold		30.03	67.15	81.94	88.30	26.54	66.70	84.18	91.12
BASE	BARThez (ours)	10.93	34.03	47.97	56.80	16.69	51.70	72.05	82.49
LARGE	C2C	39.43	79.12	92.04	96.28	33.82	75.74	91.77	96.71
	mBART	13.40	38.94	53.69	62.61	16.89	52.28	73.12	82.74
	mBARThez (ours)	15.48	43.34	58.53	67.30	17.79	53.41	73.38	82.94

Table 7: Proportion of novel n-grams in the generated summaries. C2C stands for CamemBERT2CamemBERT. Note that C2C’s high scores are misleading as many of the introduced words are irrelevant.

		CLS-books	CLS-DVD	CLS-music	PAWSX	XNLI
BASE	mBERT [†] (Devlin et al., 2019)	86.15	89.90	86.65	89.30	76.9
	CamemBERT _{BASE} [†] (Martin et al., 2020)	92.30	93.00	94.85	90.14	81.20
	FlauBERT _{BASE} [†] (Le et al., 2020)	92.30	92.45	94.10	89.49	80.60
	BARThez (ours)	94.47 _{0.17}	93.17 _{0.40}	94.97 _{0.25}	88.90 _{0.24}	80.73 _{0.40}
	BART-random	76.37 _{0.34}	73.20 _{0.65}	76.00 _{1.28}	55.27 _{0.33}	60.43 _{0.87}
LARGE	Camembert _{LARGE} (Martin et al., 2020)	95.47 _{0.33}	95.37 _{0.07}	96.00 _{0.29}	91.83 _{0.54}	85.33 _{0.05}
	Flaubert _{LARGE} [†] (Le et al., 2020)	95.00	94.10	95.85	89.34	83.40
	mBART (Liu et al., 2020)	93.40 _{0.22}	93.10 _{0.20}	93.13 _{0.79}	89.70 _{0.22}	81.07 _{0.38}
	mBARThez (ours)	94.63 _{0.05}	94.03 _{0.09}	95.30 _{0.16}	90.90 _{0.22}	81.87 _{0.50}

Table 8: Accuracy on discriminative tasks (standard deviation as subscript). [†] are taken from Le et al. (2020).

		Length	Repetitions (%)
ABSTRACT	Gold	32.12	11.47
	mBART	28.20	7.47
	mBARThez	29.45	8.60
	BARThez	29.10	14.47
	C2C	30.68	23.00
TITLE	Gold	11.42	0.93
	mBART	10.79	1.73
	mBARThez	11.03	2.27
	BARThez	11.19	2.73
	C2C	11.23	19.53

Table 9: Summary statistics.

System		Score
Gold		14.29
BASE	BARThez (ours)	21.43
LARGE	CamemBERT2CamemBERT	-75.00
	mBART	11.90
	mBARThez (ours)	27.38

Table 10: Human evaluation using Best-Worst Scaling.

evaluation scores. A manual observation revealed that actually, many of the new words introduced by C2C are irrelevant (see summary examples in appendix).

Also, like in Rothe et al. (2020), we computed the length of the summaries, and the percentage of summaries with at least one non-stopword repetition. We used as stopwords the 500 most frequent words from the system and gold summaries, across all documents. As can be seen in Table 9, for both the Abstract and Title tasks, all models generated summaries of sizes very close to that of the Gold summaries.

In terms of repetitions, the less redundant models, closest to the ground truth, are mBART and mBARThez. This is especially apparent on the Abstract task, where potential for repetition is greater.

On this task, mBART and mBARThez show less than 9% repetitions, compared with 14.5 and 23 for BARThez and C2C (resp.), and 11.5 in the references. C2C is also way more redundant than the other models and far from the reference on the Title task, with 19.5% of repetitions.

6.2 Discriminative tasks

In addition to generative tasks, BART-like models can perform discriminative tasks (Lewis et al., 2020). In the case of sequence classification, the input sequence is fed to both the encoder and the decoder, and the representation of the last token in the sequence is used by adding a classification head on top of it. When the input consists of several sentences, these sentences are separated with a special token and pasted together. We evaluate the different models on five discriminative tasks from the FLUE benchmark¹¹ (Le et al., 2020), the French equivalent of GLUE (Wang et al., 2019b).

- **CLS.** The Cross-lingual Sentiment analysis dataset (Prettenhofer and Stein, 2010) is made of Amazon reviews to be classified as positive or negative. It contains 3 product categories: books, DVD and music. The train and test sets are balanced and contain 2000 examples (each) per product category. Following Le et al. (2020), we used 20% of the train set as validation set.

- **PAWSX.** The Cross-lingual Adversarial Dataset for Paraphrase Identification (Yang et al., 2019) contains pairs of sentences, and the task is to predict whether they are semantically equivalent.

¹¹ <https://github.com/getalp/Flaubert/tree/master/flue>

There are 49401 examples for training, 1992 for development, and 1985 for testing.

- **XNLI.** The Cross-lingual NLI corpus (Conneau et al., 2018) contains pairs of sentences, and the task is to predict whether the first one (premise) entails the second one (hypothesis), contradicts it, or neither entails nor contradicts it (neutral relationship). 392702 pairs are used for training, 2490 for development, and 5010 for testing.

Training details. In all experiments, we finetuned the model for 10 epochs with a learning rate chosen from $\{10^{-4}, 5 \cdot 10^{-5}, 10^{-5}\}$ based on the best validation score. We repeated each experiment 3 times with different seeds and report the mean and standard deviation.

Results. As shown in Table 8, among the models having a BASE architecture, BARThez is best in the three sentiment analysis tasks, while being very close to CamemBERT and FlauBERT in the paraphrasing and inference tasks.

Among the LARGE models, mBARThez outperforms mBART in all tasks, showing again the importance of language-adaptive pretraining. On the other hand, CamemBERT and FlauBERT outperform mBARThez in most of the tasks, which could be attributed to the fact that CamemBERT and FlauBERT were trained for approximately 10 times more GPU hours on a monolingual French corpus. Nevertheless, Given that huge difference in monolingual training time, it is remarkable that mBARThez is so close, and sometimes outperforms, FlauBERT, with e.g., a 1.56 margin on PAWSX.

We can conclude that the ability of BARThez and mBARThez to perform well on generative tasks does not appear to come at the expense of a decrease in performance on discriminative tasks, which is in line with the results presented in the BART paper (Lewis et al., 2020).

7 Conclusion

We released BARThez and mBARThez, the first French large-scale pretrained seq2seq models, as well as a novel French summarization dataset, OrangeSum, inspired by the XSum dataset. We showed our proposed models to be very competitive with the state of the art on generative and discriminative tasks. By evaluating our models on the summarization dataset we showed that: (1) BARThez is on par with mBART while having four times less parameters, and that (2) mBARThez pro-

vides a significant boost over mBART by simply adding a relatively affordable language-adaptive phase to the pretraining. In addition, we evaluated BARThez and mBARThez on 5 sentiment analysis, paraphrasing, and natural language inference tasks against cutting edge BERT-based French language models (FlauBERT and CamemBERT), and obtained very competitive results. An interesting area for future work is to further explore language-adaptive pretraining.

Acknowledgments

We are thankful to the National Center for Scientific Research (CNRS) for giving us access to their Jean Zay supercomputer, under allocation 2020-AD011011499. Finally, We thank the three anonymous reviewers for their constructive feedback.

References

- Wissam Antoun, Fady Baly, and Hazem Hajj. 2020. [AraBERT: Transformer-based model for Arabic language understanding](#). In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, pages 9–15, Marseille, France. European Language Resource Association.
- ATILF and CLLE. 2020. [Corpus journalistique issu de l’est républicain](#). ORTOLANG (Open Resources and TOols for LANGuage) –www.ortolang.fr.
- Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. [Enriching word vectors with subword information](#). *Transactions of the Association for Computational Linguistics*, 5:135–146.
- José Cañete, Gabriel Chaperon, Rodrigo Fuentes, and Jorge Pérez. 2020. Spanish pre-trained bert model and evaluation data. In *to appear in PMLADC 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](#). In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 7057–7067.

- 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.
- Wietse de Vries, Andreas van Cranenburgh, Arianna Bisazza, Tommaso Caselli, Gertjan van Noord, and Malvina Nissim. 2019. [Bertje: A dutch bert model](#). *ArXiv preprint*, abs/1912.09582.
- Pieter Delobelle, Thomas Winters, and Bettina Berendt. 2020. [RobBERT: a Dutch RoBERTa-based Language Model](#). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3255–3265, Online. 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.
- Andreas Eisele and Yu Chen. 2010. [MultiUN: A multilingual corpus from united nation documents](#). In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*, Valletta, Malta. European Language Resources Association (ELRA).
- Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. [Don’t stop pretraining: Adapt language models to domains and tasks](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8342–8360, Online. Association for Computational Linguistics.
- Karl Moritz Hermann, Tomáš 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.
- Jeremy Howard and Sebastian Ruder. 2018. [Universal language model fine-tuning for text classification](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 328–339, Melbourne, Australia. Association for Computational Linguistics.
- Diederik P. Kingma and Jimmy Ba. 2015. [Adam: A method for stochastic optimization](#). In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. [Imagenet classification with deep convolutional neural networks](#). In *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States*, pages 1106–1114.
- 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.
- Yuri Kuratov and Mikhail Arkhipov. 2019. [Adaptation of deep bidirectional multilingual transformers for russian language](#). *ArXiv preprint*, abs/1905.07213.
- Hang Le, Loïc Vial, Jibril Frej, Vincent Segonne, Maximin Coavoux, Benjamin Lecouteux, Alexandre Al-lauzen, Benoit Crabbé, Laurent Besacier, and Didier Schwab. 2020. [FlauBERT: Unsupervised language model pre-training for French](#). In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 2479–2490, Marseille, France. European Language Resources Association.
- 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.
- Xian Li, Paul Michel, Antonios Anastasopoulos, Yonatan Belinkov, Nadir Durrani, Orhan Firat, Philipp Koehn, Graham Neubig, Juan Pino, and Hassan Sajjad. 2019. [Findings of the first shared task on machine translation robustness](#). In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 91–102, Florence, Italy. 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.
- Yang Liu and Mirella Lapata. 2019. [Text summarization with pretrained encoders](#). 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 3730–3740, Hong Kong, 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.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. [Roberta: A robustly optimized bert pretraining approach](#). *ArXiv preprint*, abs/1907.11692.
- Jordan J Louviere, Terry N Flynn, and Anthony Alfred John 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*, pages 7203–7219, Online. Association for Computational Linguistics.
- Tomás Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. [Distributed representations of words and phrases and their compositionality](#). In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 3111–3119.
- 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.
- 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.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. [GloVe: Global vectors for word representation](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.
- Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. [Deep contextualized word representations](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.
- Marco Polignano, Pierpaolo Basile, Marco de Gemmis, Giovanni Semeraro, and Valerio Basile. 2019. [Alberto: Italian bert language understanding model for nlp challenging tasks based on tweets](#). In *CLiC-it*.
- Peter Prettenhofer and Benno Stein. 2010. [Cross-language text classification using structural correspondence learning](#). In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1118–1127, Uppsala, Sweden. Association for Computational Linguistics.
- Weizhen Qi, Yu Yan, Yeyun Gong, Dayiheng Liu, Nan Duan, Jiusheng Chen, Ruofei Zhang, and Ming Zhou. 2020. [ProphetNet: Predicting future n-gram for sequence-to-SequencePre-training](#). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2401–2410, Online. Association for Computational Linguistics.
- Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.
- Anna Rogers, Olga Kovaleva, and Anna Rumshisky. 2020. [A primer in BERTology: What we know about how BERT works](#). *Transactions of the Association for Computational Linguistics*, 8:842–866.
- Sascha Rothe, Shashi Narayan, and Aliaksei Severyn. 2020. [Leveraging pre-trained checkpoints for sequence generation tasks](#). *Transactions of the Association for Computational Linguistics*, 8:264–280.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. [Neural machine translation of rare words with subword units](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany. Association for Computational Linguistics.
- Raivis Skadiņš, Jörg Tiedemann, Roberts Rozis, and Daiga Dekšne. 2014. [Billions of parallel words for free: Building and using the EU bookshop corpus](#). In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, pages 1850–1855, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2019. [MASS: masked sequence to sequence pre-training for language generation](#). In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019*,

Long Beach, California, USA, volume 97 of *Proceedings of Machine Learning Research*, pages 5926–5936. PMLR.

Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. 2019. [Portuguese named entity recognition using bert-crf](#). *ArXiv preprint*, abs/1909.10649.

Jörg Tiedemann. 2012. [Parallel data, tools and interfaces in OPUS](#). In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2214–2218, Istanbul, Turkey. European Language Resources Association (ELRA).

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.

Alex Wang, Yada Pruksachatkun, Nikita Nangia, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. 2019a. [Superglue: A stickier benchmark for general-purpose language understanding systems](#). In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 3261–3275.

Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. 2019b. [GLUE: A multi-task benchmark and analysis platform for natural language understanding](#). In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

Yinfei Yang, Yuan Zhang, Chris Tar, and Jason Baldridge. 2019. [PAWS-X: A cross-lingual adversarial dataset for paraphrase identification](#). 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 3687–3692, Hong Kong, China. Association for Computational Linguistics.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. [Bertscore: Evaluating text generation with BERT](#). In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

Appendices

In what follows, we provide, for 10 documents randomly selected from OrangeSum’s test set, the reference and model summaries for each task (Abstract and Title).

	Document	"Nous pouvons confirmer à ce stade que cinq personnes ont péri. Au moins 70 personnes sont dans un état inconscient dans un hôpital non loin, et dans l'ensemble, entre 200 et 500 personnes reçoivent des soins", a déclaré Swaroop Rani, une responsable de la police de Visakhapatnam, dans l'État de l'Andhra Pradesh. Opérée par LG Polymers, l'usine est située en bordure de la ville industrielle et portuaire de Visakhapatnam. L'agglomération compte une population d'environ 5 millions de personnes. Le gaz "avait été laissé là à cause du confinement. Cela a mené à une réaction chimique et de la chaleur est apparue dans les réservoirs, et le gaz a fui à cause de cela", a expliqué Mme Rani. L'Inde est actuellement confinée depuis fin mars pour lutter contre la pandémie de coronavirus. "Nous avons reçu un appel d'urgence de villageois autour de 03H30 du matin aujourd'hui (mercredi 22H00 GMT, ndlr). Ils disaient qu'il y avait du gaz dans l'air", a-t-elle poursuivi. "Nous sommes arrivés immédiatement sur place. On pouvait sentir le gaz dans l'air et il ne n'était possible à aucun d'entre nous de rester là plus de quelques minutes", a-t-elle décrit. L'Inde a été le théâtre en décembre 1984 d'un des pires accidents industriels de l'Histoire, lorsque 40 tonnes de gaz s'étaient échappées d'une usine de pesticides de la ville de Bhopal (centre). Quelque 3.500 personnes avaient péri en quelques jours, principalement dans des bidonvilles situés autour de cette usine d'Union Carbide, et des milliers d'autres dans les années qui ont suivi.
ABSTRACT	Gold	Au moins cinq personnes ont péri et plus d'un millier ont été hospitalisées à la suite d'une fuite de gaz dans une usine chimique du sud-est de l'Inde, ont annoncé jeudi la police et une source officielle.
	mBART	Cinq personnes sont mortes mercredi dans l'accident d'une usine de pesticides qui s'est produite en Inde, à la suite du confinement lié à l'épidémie de Covid-19, ont indiqué des responsables.
	mBARThez	Une explosion a fait cinq morts mercredi dans une usine de pesticides du centre de l'Inde, après que le gaz a fui dans les réservoirs après une réaction chimique, ont indiqué les autorités.
	BARThez	Une dizaine de personnes ont péri et des centaines d'autres ont été blessées mercredi dans une usine de pesticides près de Visakhapatnam, dans le sud de l'Inde, a annoncé la police.
	C2C	Au moins vingt personnes sont mortes, dont cinq sont mortes et cinq sont portées disparues, selon un bilan officiel lundi après-midi en Inde, faisant craindre une fuite de gaz meurtrière dans le pays, selon une source gouvernementale à l'AFP.
TITLE	Gold	Fuite de gaz dans une usine en Inde: 5 morts, au moins 1.000 personnes hospitalisées
	mBART	Inde: cinq morts dans un accident de la usine de pesticides
	mBARThez	Inde: au moins cinq morts dans le crash d'une usine de pesticides
	BARThez	Inde: cinq morts dans un glissement de terrain à Visakhapatnam
	C2C	Inde: cinq morts dans un gaz mortel dans un usine de recyclage

Table 11: C2C stands for CamemBERT2CamemBERT. OrangeSum document 12158.

Document	<p>De nombreux scientifiques occidentaux ont fait part de leurs doutes quant à la rapidité avec laquelle ce vaccin aurait été mis au point. Le ministre américain de la Santé Alex Azar s'est fait l'écho mercredi de leurs points de vue, à l'issue d'une visite de trois jours à Taïwan. "Il est important que nous fournissions des vaccins sans danger et efficaces et que les données soient transparentes... Ce n'est pas une course pour être le premier", a-t-il déclaré à la presse lors d'une conférence téléphonique. "Je dois souligner que deux des six vaccins américains dans lesquels nous avons investi sont entrés dans la phase des essais cliniques il y a trois semaines, alors que le vaccin russe ne fait que commencer", a-t-il ajouté. "Les données des premiers essais en Russie n'ont pas été divulguées, ce n'est pas transparent", a estimé le ministre américain. Mardi, le président russe Vladimir Poutine a annoncé le développement par son pays du "premier" vaccin sans danger contre le Covid-19, affirmant que l'une de ses filles se l'est fait inoculer. Ce vaccin a été baptisé "Sputnik V" (V comme vaccin, ndlr), en référence à la victoire politico-scientifique russe qu'était la mise en orbite en 1957 du satellite du même nom en pleine Guerre froide. Peu après la déclaration du Kremlin, l'Organisation mondiale de la santé (OMS) a réagi en appelant à la prudence, rappelant que la "pré-qualification" et l'homologation d'un vaccin passaient par des procédures "rigoureuses". De nombreux pays se sont lancés dans une véritable* course contre la montre pour trouver un vaccin efficace contre le coronavirus qui a tué plus de 740.000 personnes à travers la planète depuis son apparition l'an dernier en Chine. Les Etats-Unis sont le pays le touché avec 5,1 millions de cas de Covid-19 qui a fait plus de 164.000 morts. M. Azar s'est dit confiant sur la capacité des Américains à mettre au point un vaccin. "Nous pensons qu'il est très crédible que nous ayons des dizaines de millions de doses de vaccin de référence, sûres et efficaces d'ici la fin de cette année, et plusieurs centaines de millions de doses au début de l'année prochaine", a-t-il affirmé. Le président américain a lancé l'opération "Warp Speed" qui vise explicitement à obtenir de quoi vacciner tous les Américains dès janvier 2021.</p>
ABSTRACT	<p>Gold Le ministre américain de la Santé a fait part mercredi de son scepticisme après l'annonce du développement par la Russie du "premier" vaccin contre le coronavirus assurant une "immunité durable".</p> <p>mBART Le ministre américain de la Santé s'est exprimé mercredi sur la possibilité d'un vaccin efficace contre le coronavirus.</p> <p>mBARThez Le ministre de la Santé américain Alex Azar a déclaré mercredi que les Etats-Unis et la Russie ont fait des essais cliniques d'un vaccin contre le coronavirus "sans danger", alors que le président russe Vladimir Poutine a déclaré mardi avoir inoculé une de ses filles.</p> <p>BARThez Les Etats-Unis sont le pays le plus touché par la pandémie de coronavirus après la Russie qui a pourtant annoncé avoir mis au point le "premier vaccin", a estimé le ministre américain de la Santé.</p> <p>C2C Le ministre américain de la Santé a souligné que la Russie avait développé des capacités capables de détecter et de tester si le pays n'était pas vaccin contre le nouveau coronavirus, mais a jugé prudent "dans l'attente de la publication d'une réponse scientifique",</p>
TITLE	<p>Gold Coronavirus: le ministre américain de la Santé sceptique au sujet du vaccin russe</p> <p>mBART Vaccin " sans danger": les Américains investis en Russie, selon Alex Azar</p> <p>mBARThez Vaccin russe: les Américains appelés à la prudence</p> <p>BARThez Un vaccin russe contre le Covid-19 en vue aux Etats-Unis, selon le ministre américain de la Santé</p> <p>C2C Coronavirus: les Etats-Unis pas en "cours de combattant" face à un vaccin expérimental</p>

Table 12: C2C stands for CamemBERT2CamemBERT. OrangeSum document 33555.

Document	<p>Une première depuis la Seconde guerre mondiale, la consommation d'alcool ne baisse plus en France. L'Académie nationale de médecine a appelé lundi 29 avril les pouvoirs publics à "prendre des mesures plus fortes" pour lutter contre les problèmes de santé publique causés par la consommation d'alcool. "Pour la première fois depuis la Seconde guerre mondiale, la consommation d'alcool ne baisse plus en France. C'est une défaite majeure pour la santé publique, car l'alcool en est un déterminant fondamental", estime l'Académie dans un communiqué diffusé lundi 29 avril. L'organisme déplore en particulier "l'affaiblissement continu de la loi Evin sous la pression du lobby alcoolier, jusqu'à autoriser la publicité sur l'internet, support médiatique particulièrement affectueux des jeunes". L'alcool serait la première cause évitable* de mortalité des 15-30 ans, selon l'Académie de Médecine. Elle invite donc le gouvernement à revenir aux "principes initiaux" de la loi. Pour un pictogramme plus visible pour les femmes enceintes À l'instar d'autres institutions et associations, l'Académie recommande d'interdire la publicité pour l'alcool et de faire figurer sur les boissons alcoolisées la mention "l'alcool est dangereux pour la santé" (et non le seul excès). L'Académie de médecine veut également voir taxées les boissons au gramme d'alcool et demande la mise en place d'un prix minimum de vente par gramme d'alcool, comme c'est le cas en Ecosse depuis un an. Elle réclame également un pictogramme plus grand et plus lisible sur les bouteilles pour "dissuader de toute consommation la femme enceinte ou qui désire l'être". L'académie de médecine pointe clairement la responsabilité du lobby alcoolier. "Malgré l'enjeu de prévenir la première cause de retard mental évitable* du nouveau-né et de l'enfant, les discussions pour l'agrandir et le contraster s'enlisent depuis des années face à l'opposition farouche du lobby alcoolier". L'alcool serait la première cause de retard mental de l'enfant et de démence précoce souligne l'organisme. Un quart des Français boit trop Dans des chiffres publiés mi-février, Santé publique France avait indiqué que la consommation des Français n'avait quasiment pas reculé depuis 10 ans, passant de 27 g à 26 g d'alcool pur par jour entre 2009 et 2015. "C'est en février 2019 que Santé Publique France annonce que la consommation française d'alcool est la même en 2017 qu'en 2013", note l'académie dans son communiqué. Près d'un quart des Français, soit environ 10,5 millions d'adultes, boivent trop d'alcool, avait également estimé fin mars Santé publique France. L'agence sanitaire a diffusé de nouveaux repères de consommation, résumés par le message "pour votre santé, c'est maximum deux verres par jour, et pas tous les jours". L'alcool constitue la deuxième cause de mortalité évitable* après le tabac, avec 41.000 décès qui lui sont attribuables chaque année en France, 30.000 hommes et 11.000 femmes. L'alcool "est impliqué dans 40% des violences faites aux femmes et aux enfants et un tiers des décès par accidents de la route", ajoute l'Académie dans son communiqué.</p>
<p>ABSTRACT</p> <p>Gold</p> <p>mBART</p> <p>mBARThez</p> <p>BARThez</p> <p>C2C</p>	<p>Elle demande des "mesures plus fortes" pour lutter contre les problèmes de santé causés en France par une consommation d'alcool qui ne diminue plus.</p> <p>En février 2019, Santé publique France avait indiqué que la consommation des Français n'avait quasiment pas reculé depuis 10 ans.</p> <p>Près d'un quart des Français boivent trop d'alcool.</p> <p>L'Académie de médecine réclame notamment "l'affaiblissement continu de la loi Evin sous la pression du lobby alcoolier", jusqu'à autoriser la publicité sur l'internet.</p> <p>À l'inverse de ce qui se fait en France, la mesure doit inciter à la consommation d'alcool dès l'âge de 18 ans.</p>
<p>TITLE</p> <p>Gold</p> <p>mBART</p> <p>mBARThez</p> <p>BARThez</p> <p>C2C</p>	<p>Stagnation de la consommation d'alcool en France : "une défaite majeure pour la santé publique"</p> <p>Santé : l'Académie de médecine demande des mesures plus fortes</p> <p>alcool : l'Académie de médecine appelle le gouvernement à des mesures plus fortes</p> <p>Alcool : une "défaite majeure" pour la santé publique, selon l'Académie de médecine</p> <p>La consommation d'alcool en forte hausse : l'Académie de médecine appelle à plus de fermeté</p>

Table 13: C2C stands for CamemBERT2CamemBERT. OrangeSum document 25148.

Document	<p>De petites dimensions (20 cm de largeur et 30 de hauteur), ces ouvertures à hauteur d'homme percées à côté du porche des somptueux palais appartenant aux grandes familles florentines servaient à écouler le vin directement du producteur au consommateur. Au fil des siècles, ce détail architectural et sa fonction sont tombés dans les oubliettes de l'Histoire jusqu'à ce que Massimo Casprini, un érudit florentin, parte à leur redécouverte et y consacre un livre, "I finestrini del vino" ("Les fenêtres à vin"), publié en 2005. Ces fenêtres "ont été créées à partir de 1532 après la chute de la République, quand les Médicis sont revenus au pouvoir et ont voulu favoriser l'agriculture, incitant les grands propriétaires florentins à investir dans les oliveraies et les vignes (...) tout en leur donnant des avantages fiscaux pour revendre directement leur production en ville", explique à l'AFP M. Casprini lors d'une promenade à travers les rues de Florence dans la touffeur estivale. Unique restriction: "Ils pouvaient y vendre seulement le vin de leur propre production et sous un format particulier d'environ 1,4 litre." L'autre fonction de ces petites fenêtres était sociale, en permettant aux gens du peuple d'acquérir du vin à prix plus raisonnable que chez les commerçants, sans intermédiaire", ajoute-t-il, précisant dans un sourire qu'"à l'époque la consommation de vin était énorme". - Episodes de peste - A l'heure du coronavirus et de la distanciation sociale, Massimo Casprini rappelle que "grâce à ce système on évitait les contacts", alors qu'"épidémies et épisodes de peste étaient très fréquents au XVIe siècle". "En effet, la fenêtre à vin était fermée par un panneau de bois, le client se présentait et frappait avec le heurtoir, à l'intérieur il y avait un caviste qui prenait la bouteille vide et la remplissait. Il n'y avait donc pas de contact direct!" s'extasie le fringant septuagénaire, également amateur de motos anciennes et auteur de quelque 70 ouvrages centrés sur la capitale toscane. Jusqu'ici, 267 de ces fenêtres à vin ont été répertoriées en Toscane, dont 149 dans le centre de Florence. "Il y en avait beaucoup plus!" estime M. Casprini, "presque tous les propriétaires terriens avaient une fenêtre à vin, mais nombre d'entre elles ont disparu, notamment lors des bombardements de la Seconde Guerre mondiale". Certaines ont aussi été murées, mais grâce à l'œil de lynx de notre expert on réussit encore à reconnaître les contours de leur encadrement en pietra serena (grès gris) ou pierre des carrières de Fiesole, près de Florence. Dans le fil du livre du professeur Casprini a été fondée une association, baptisée "Le buchette del vino", qui recense et appose une plaque sur chaque fenêtre. Son site internet (https://buchettedelvino.org/) propose même une carte interactive permettant de partir à leur découverte, ainsi qu'une galerie de photos et une présentation historique de ces petits trésors architecturaux. On y cite par exemple un guide en français de Florence datant de 1892 qui mentionne la fenêtre d'un palais: "cette cave assez renommée pour ses vins millésimés ne livre aux consommateurs que ceux provenant des propriétés de la marquise Leonia degli Albizi Frescobaldi". Tombées en désuétude, les "finestrini del vino" font aujourd'hui l'objet d'un regain d'intérêt et d'une forme de recyclage: présentoir de magasin, passe-plat dans un café, ou encore petit autel dédié à la Vierge. Même si elles sont protégées par la loi, M. Casprini déplore que "trois fenêtres ont déjà disparu" depuis son premier recensement en 2005.</p>
<p style="writing-mode: vertical-rl; transform: rotate(180deg);">ABSTRACT</p> <p>Gold</p> <p>mBART</p> <p>mBARThez</p> <p>BARThez</p> <p>C2C</p>	<p>Florence, joyau de la Renaissance, peut s'enorgueillir d'un patrimoine mondialement célèbre, mais dont certains détails restent encore aujourd'hui méconnus: c'est le cas des discrètes "fenêtres à vin" ornant la façade de certains palais, qui permettaient la vente de vin "sans contact", un concept redevenu d'actualité en ces temps de coronavirus.</p> <p>"A l'heure du coronavirus et de la distanciation sociale, il n'y avait pas de contact direct !" A Florence, des fenêtres à vin, remplacées par des pierres, auraient été oubliées dans les oubliettes de l'Histoire.</p> <p>Massimo Casprini, spécialiste des fenêtres à vins, est revenu mercredi sur la révolution de Florence (Italie) où il a redécouvert l'existence de ces ouvertures en plein air et à ciel ouvert qui permettaient, autrefois, à des propriétaires de vins de revendre leur production à la ville.</p> <p>De 1532 à nos jours, les fenêtres à vin des palais anciens de Florence sont les plus souvent murées, un détail qui a sans doute survécu à l'épidémie de nouveau coronavirus.</p> <p>Au lieu de la pandémie de coronavirus, un jardin italien a retrouvé des crus du monde entier: ils étaient des caves à vin français, à quelques dizaines de mètres du sol, pour ne pas être contaminés par le Covid-19.</p>
<p style="writing-mode: vertical-rl; transform: rotate(180deg);">TITLE</p> <p>Gold</p> <p>mBART</p> <p>mBARThez</p> <p>BARThez</p> <p>C2C</p>	<p>Virus: comment la Florence des Médicis inventa la vente de vin "sans contact"</p> <p>Les fenêtres à vin sont tombées dans les oubliettes de l'Histoire</p> <p>Florence: les fenêtres à vin cachées dans un livre</p> <p>Florence: des fenêtres à vin traditionnelles à l'heure des peste</p> <p>Les fenêtres de la Florence en "huile de vin" : actualité automobile, infos, scoop</p>

Table 14: C2C stands for CamemBERT2CamemBERT. OrangeSum document 34657.

Document	<p>L'ancien chef de l'État était entendu depuis mardi matin, avec une interruption dans la nuit, dans les locaux de l'office anticorruption (OCLCIPP) situés à Nanterre (Hauts-de-Seine). L'ancien président de l'UMP a regagné son domicile parisien du XVI^e arrondissement après la fin de sa garde à vue. Également entendu, mais sous le statut de "suspect libre", Brice Hortefeux, un proche de l'ex-président qui occupa plusieurs postes ministériels pendant le quinquennat Sarkozy (2007-2012), a de son côté quitté les locaux de l'office anticorruption mardi soir, assurant sur Twitter avoir apporté des précisions pour "permettre de clore une succession d'erreurs et de mensonges". Depuis la publication, en mai 2012, par le site d'informations Mediapart d'un document libyen - attribué à l'ex-chef des renseignements Moussa Koussa - accréditant un financement d'environ 50 millions d'euros, les investigations des juges ont considérablement avancé. Plusieurs protagonistes du dossier, dont plusieurs ex-responsables libyens, ont accrédité la thèse de versements illicites. Ziad Takieddine persiste et signeLe sulfureux homme d'affaires Ziad Takieddine a lui-même assuré avoir remis entre fin 2006 et début 2007 trois valises contenant 5 millions d'euros en provenance du régime de Kadhafi à Nicolas Sarkozy, alors ministre de l'Intérieur, et à son directeur de cabinet Claude Guéant. Sur BFMTV, il a réédité ses accusations mais répété que cet argent "n'était pas lié à la campagne présidentielle" de 2007. "Cet argent faisait partie des accords entre les deux pays sur le contrôle des frontières maritimes, avec échanges d'informations", a précisé l'homme d'affaires, mis en examen autour de ce dossier pour complicité de corruption et complicité de diffamation. "Il y avait un devoir de former en France des équipes libyennes avant la livraison du matériel. Dans ce cadre-là, il y avait des formations à destination de quelques centaines de Libyens. Ils ont établi en France que ça allait coûter dans les cinq millions d'euros", a-t-il ajouté.Nicolas Sarkozy "est un vrai menteur et vous allez voir, il va passer son temps avec les juges d'instruction à dire 'non, non, non c'est pas vrai'. Tout ça pour gagner du temps, c'est sa méthode habituelle", a également lancé Ziad Takieddine sur BFMTV mercredi, assurant "dire la vérité". L'ancien chef de l'État a toujours rejeté ces mises en cause. D'autres dignitaires libyens ont démenti tout financement de la Libye de Mouammar Kadhafi, que Nicolas Sarkozy avait reçu en grande pompe à l'Élysée en 2007. De nouveaux éléments compromettantsOuverte notamment pour "détournements de fonds publics" et "corruption active et passive", l'enquête a été élargie en janvier à des soupçons de "financement illégal de campagne électorale", suite à un rapport de l'office anticorruption qui pointe une circulation importante d'argent liquide dans l'entourage de Nicolas Sarkozy durant la campagne 2007. Selon Le Monde, plusieurs anciens dignitaires du régime Kadhafi auraient livré de récents témoignages confirmant les soupçons de financement illicite. Les investigations ont aussi mis en lumière un virement de 500.000 euros perçu par Claude Guéant en mars 2008, en provenance d'une société d'un avocat malaisien. L'ex-secrétaire général de l'Élysée a toujours affirmé qu'il s'agissait du fruit de la vente de deux tableaux, sans convaincre les juges qui l'ont mis en examen notamment pour "blanchiment de fraude fiscale en bande organisée". Les juges s'interrogent également sur la vente suspecte en 2009 d'une villa située à Mougins, sur la Côte d'Azur, à un fonds libyen. Ils soupçonnent l'homme d'affaires Alexandre Djouhri d'avoir été le véritable* propriétaire de ce bien et de l'avoir cédé pour 10 millions d'euros, soit plus du double du prix du marché, ce qui aurait pu permettre de dissimuler d'éventuels versements occultes.</p>
<p style="writing-mode: vertical-rl; transform: rotate(180deg);">ABSTRACT</p> <p style="text-align: center;">Gold</p> <p style="text-align: center;">mBART</p> <p style="text-align: center;">mBARThez</p> <p style="text-align: center;">BARThez</p> <p style="text-align: center;">C2C</p>	<p>Après une vingtaine d'heures, la garde à vue de Nicolas Sarkozy s'est achevée mercredi soir. L'ancien président a été mis en examen pour "corruption passive", "financement illégal de campagne électorale" et "recel de fonds publics libyens" et placé sous contrôle judiciaire dans le cadre de l'enquête sur des soupçons de financement de sa campagne présidentielle de 2007 par la Libye de Mouammar Kadhafi.</p> <p>Ziad Takieddine, mis en examen autour de l'affaire des soupçons de financement libyen de la campagne présidentielle de 2007 de Nicolas Sarkozy, a de nouveau quitté les locaux de son établissement, à Nanterre, dans la nuit de mardi à mercredi.</p> <p>Cinq jours après la révélation d'un document par Mediapart, l'ancien président de l'UMP et principal suspect dans l'affaire des soupçons de financement libyen de sa campagne présidentielle de 2007 a quitté mardi soir les locaux où il était auditionné. Brice Hortefeux et Claude Guéant ont apporté des précisions.</p> <p>L'ancien président de la République Nicolas Sarkozy a quitté mardi matin les locaux de l'office anticorruption où il était entendu. Les soupçons de financement libyen de sa campagne présidentielle de 2007.</p> <p>Nicolas Sarkozy est mis en examen dans le cadre de l'enquête sur les soupçons de financement libyen de sa campagne présidentielle de 2007. Selon plusieurs médias, l'ancien chargé de mission a dit mercredi n'être "pas au courant" de ce que l'ex-</p>
<p style="writing-mode: vertical-rl; transform: rotate(180deg);">TITLE</p> <p style="text-align: center;">Gold</p> <p style="text-align: center;">mBART</p> <p style="text-align: center;">mBARThez</p> <p style="text-align: center;">BARThez</p> <p style="text-align: center;">C2C</p>	<p>Soupçons de financement libyen : Nicolas Sarkozy mis en examen</p> <p>Affaire libyenne : Nicolas Sarkozy en garde à vue</p> <p>Affaire libyenne : Nicolas Sarkozy entendu par les juges</p> <p>Nicolas Sarkozy en garde à vue, la piste d'un financement libyen s'éloigne</p> <p>Nicolas Sarkozy est "un vrai traître" selon l'entourage de Nicolas Sarkozy</p>

Table 15: C2C stands for CamemBERT2CamemBERT. OrangeSum document 22208.

Document	<p>Jean-Paul Dufrègne a passé un sale quart d'heure sur les réseaux sociaux mercredi soir. Cet élu communiste de l'Allier a été filmé par les caméras de TF1, dans un reportage diffusé le 4 avril au journal de 20 heures. Mais téléspectateurs et internautes n'ont nullement prêté attention aux arguments du député sur les inquiétudes persistantes des territoires ruraux et la réforme institutionnelle sur laquelle planche le gouvernement. Non, ils étaient bien trop captivés par son compteur de vitesse, filmé le temps de quelques plans par les caméras de la première chaîne, comme le relève LCI. Car, sur une route départementale limitée à 90 km/heure, Jean-Paul Dufrègne avait le pied au plancher. Son compteur affichait 124 km/heure, plus de 30 km/heure au-dessus de la limite autorisée. Une infraction que n'ont pas manqué de relever de nombreux internautes. " Trois points et 135 euros d'amende ", note un utilisateur de Twitter. " Bonjour, les limitations de vitesse ne s'appliquent pas aux parlementaires ? ", ironise un autre. Opposant au 80km/heure Certains ont par ailleurs fait le lien entre les positions politiques de l'élú communiste et cet excès de vitesse. Car Jean-Paul Dufrègne est un farouche opposant au projet du gouvernement de limiter le réseau français de routes secondaires à 80 km/heure. Avec une trentaine d'autres élus du Massif Central, il avait d'ailleurs adressé une lettre ouverte à Emmanuel Macron sur le sujet, dénonçant une mesure " injuste et pénalisante ", et un frein au développement du Massif Central.</p>
ABSTRACT	<p>Gold Jean-Paul Dufrègne, élu de l'Allier, a fait sensation sur les réseaux sociaux après être apparu dans un reportage de TF1. On le voyait rouler à plus de 120 km/heure sur une route secondaire.</p> <p>mBART L'élú communiste de l'Allier a été filmé le 4 avril au journal de 20 heures de TF1, comme le relève LCI. Son compteur affichait 124 km/heure au-dessus de la limite autorisée.</p> <p>mBARThez Son compteur affichait 124 km/heure au-dessus de la limite autorisée.</p> <p>BARThez Un élú communiste de l'Allier a été filmé le temps d'un reportage par les caméras de TF1, dans la soirée du 4 avril, alors qu'il roulait 90 km/heure au-dessus de la limite autorisée.</p> <p>C2C Un député communiste du Puy-de-Dôme n'a pas manqué de se faire remarquer en pleine séance de questions au gouvernement. Il était l'invité de Jean-Marc Remontini, l'élú apparenté du PCF, qui a fait l'éloge de</p>
TITLE	<p>Gold Un élú épinglé à 124 km/heure sur une route limitée à 90</p> <p>mBART Un élú communiste flashé à 124 km/heure sur TF1</p> <p>mBARThez Un élú communiste flashé à 124 km/heure sur TF1</p> <p>BARThez Un élú communiste passe un sale quart d'heure à cause de son compteur de vitesse</p> <p>C2C Ce député du Doubs qui n'a plus le temps de répondre aux radars</p>

Table 16: C2C stands for CamemBERT2CamemBERT. OrangeSum document 22077.

Document	<p>Mais où est donc passé Gérard Thomassin ? L'acteur français qui avait obtenu un César en 1991 est introuvable depuis le 28 août dernier, rapporte RTL. Le comédien âgé de 45 ans devait se rendre à un rendez-vous judiciaire dans une affaire de meurtre. Mais il ne s'y est jamais rendu. Et depuis, c'est toute sa famille qui s'inquiète. Interrogé par RTL, le frère de l'acteur, Jérôme Thomassin, a montré toute son inquiétude avant d'apporter des détails sur la journée du 28 août. Selon lui, Gérard Thomassin a bien "pris le train Rochefort-Lyon pour se rendre à la confrontation avec deux autres mis en examen". Parmi ces hommes, précise RTL, le principal suspect dans cette affaire de meurtre dans un bureau de poste. Les avocats du comédien qui appartiennent au cabinet d'Éric Dupond-Moretti ont signalé "une disparition inquiétante" au commissariat de Rochefort (Charente-Maritime) où l'acteur vivait. En tout état de cause, son frère était "très heureux de pouvoir se rendre à ce rendez-vous judiciaire." "L'affaire Burgod" L'affaire remonte à 2013, lorsque Gérard Thomassin est interpellé et mis en examen pour "vol avec arme et homicide sur une personne chargée d'une mission de service public". Une employée de La Poste, Catherine Burgod, enceinte, avait été tuée de 28 coups de couteau. Tenu responsable, l'acteur avait été incarcéré en 2013 avant d'être remis en liberté, mais placé sous contrôle judiciaire en octobre 2015. Sauf qu'il décide de briser son bracelet électronique et retourne en prison. Gérard Thomassin sort finalement en 2016, après trois ans de détention provisoire, la limite. L'affaire prend une autre tournure en 2017 et 2018 avec l'arrestation d'un suspect et la mise en examen d'un autre, mais la justice ne parvient toujours pas à trancher. La reconstitution du jeudi 29 août aurait dû permettre une confrontation entre les trois protagonistes, mais Gérard Thomassin ne s'est jamais présenté, au grand dam de l'avocate des parties civiles. Aujourd'hui, la famille de l'acteur demande de vraies recherches. Ses appels sonnent dans le vide. Ses proches craignent qu'il ait pu faire une mauvaise rencontre en cours de route. Le comédien avait reçu le César du Meilleur jeune espoir en 1991 pour son rôle dans "Le Petit Criminel" de Jacques Doillon, avant de connaître une carrière ponctuée de plus ou moins grands rôles.</p>
ABSTRACT	<p>Gold mBART mBARThez BARThez C2C</p> <p>Gérald Thomassin a disparu fin août selon ses proches qui s'inquiètent, révèle RTL. Il devait se rendre à un rendez-vous judiciaire pour éclairer une affaire d'homicide.</p> <p>Gérald Thomassin a disparu fin août. Il devait se rendre à un rendez-vous judiciaire pour éclairer une affaire d'homicide, révèle RTL.</p> <p>Gérald Thomassin a disparu fin août. Il devait se rendre, selon RTL, à un rendez-vous judiciaire pour éclairer une affaire de meurtre.</p> <p>Le comédien de 45 ans devait se rendre à un rendez-vous judiciaire dans une affaire de meurtre, mais il ne s'y est jamais rendu, rapporte RTL.</p> <p>Gérald Thomassin a disparu fin août. Il se rendait à un rendez-vous judiciaire pour éclairer une affaire d'homicide.</p>
TITLE	<p>Gold mBART mBARThez BARThez C2C</p> <p>Mystérieuse disparition d'un acteur césarisé soupçonné de meurtre</p> <p>Disparition de l'acteur Gérard Thomassin : sa famille s'inquiète</p> <p>L'acteur Gérard Thomassin porté disparu depuis le 28 août</p> <p>L'acteur Gérard Thomassin porté disparu depuis le 28 août</p> <p>Disparition de l'acteur Gérard Thomassin : la famille n'est plus introuvable</p>

Table 17: C2C stands for CamemBERT2CamemBERT. OrangeSum document 22168.

Document	<p>Dans un rapport adressé aux ministres de l'Intérieur, de la Justice, et à la secrétaire d'Etat à l'Egalité femmes-hommes Marlène Schiappa, les cinq députés chargés d'étudier la verbalisation du harcèlement de rue recommandent la mise en place d'"une contravention de 4e classe d'outrage sexiste et sexuel". L'infraction devra être constatée "en flagrance" par les agents de la toute récente "police de proximité du quotidien", précise leur texte, qui, selon les informations du Huffington Post, devrait être remis mercredi 28 février. Jusqu'à 1.500 euros d'amendesLe montant de l'amende forfaitaire serait de 90 euros pour un paiement immédiat, 200 euros pour un paiement sous 15 jours et 350 euros en peine majorée. En cas de circonstances aggravantes (si l'auteur est dépositaire de l'autorité publique, en cas de réunion, ou de bande organisée), une contravention de 5e classe (jusqu'à 1.500 euros) pourrait être délivrée par un tribunal de police. Pour Sophie Auconie (UDI, Agir et Indépendants), Laetitia Avia (LREM), Erwan Balanant (Modem), Elise Fajgeles (LREM) et Marietta Karamanli (Nouvelle gauche), le harcèlement subi dans l'espace public est un "fléau". Ils estiment nécessaire de "définir une nouvelle infraction visant à sanctionner, entre autres, les gestes déplacés, les sifflements, les regards insistants ou remarques obscènes, le fait de suivre volontairement à distance une personne créant ainsi une situation d'angoisse", soulignent-ils. 68% des Français favorables aux amendesLe rapport souhaite également que les auteurs participent à un stage de sensibilisation à l'égalité femmes-hommes, et que la police municipale et les agents des services de sécurité des transports soient habilités à constater cette infraction. D'après un sondage Opinionway réalisé pour Public Sénat, Les Echos et Radio Classique et publié le 5 février, une large majorité de Français est favorable à la mise en place d'une amende pénalisant le harcèlement de rue. À la question "êtes-vous favorable ou pas favorable à ce que le harcèlement de rue (sifflements, remarques...) soit passible d'une amende ?", 68 % des personnes interrogées se disent favorables (40 % "plutôt favorables" et 28 % "tout à fait favorables"). 30 % y sont opposés (23 % "plutôt opposés" et 7 % "tout à fait opposés") et 2 % ne se prononcent pas.</p>
ABSTRACT	<p>Gold Des parlementaires préconisent de créer une infraction d'"outrage sexiste" sanctionnant d'une amende immédiate de 90 euros "tout propos, comportement ou pression à caractère sexiste ou sexuel" dans l'espace public.</p> <p>mBART Selon un rapport, dévoilé par le Huffington Post, le gouvernement envisage une amende forfaitaire de 90 euros pour lutter contre le harcèlement de rue. En cas de circonstances aggravantes, elle pourrait être délivrée par un tribunal de police.</p> <p>mBARThez Dans un rapport adressé aux ministres de l'Intérieur, de la Justice et à la secrétaire d'Etat à l'Egalité femmes-hommes, les députés chargés d'étudier la verbalisation du harcèlement de rue recommandent la mise en place d'une contravention de 4e classe.</p> <p>BARThez D'après un sondage Opinionway réalisé pour Public Sénat, Les Echos et Radio Classique, une large majorité de Français sont favorables à la mise en place d'une amende pénalisant le harcèlement de rue.</p> <p>C2C Selon un sondage Elabe pour Le Huffington Post, 54% des Français sont opposés au projet de loi sur le harcèlement de rue. Une première en soi, alors que la question de l'emprise sexuelle se pose déjà : les contraventions seront en effet posées</p>
TITLE	<p>Gold Harcèlement de rue : bientôt une amende immédiate de 90 euros ?</p> <p>mBART Harcèlement de rue : vers une contravention de 4e classe ?</p> <p>mBARThez Harcèlement de rue : vers une contravention de 4e classe ?</p> <p>BARThez Harcèlement de rue : vers une contravention de 4e classe ?</p> <p>C2C Harcèlement de rue : un rapport préconise une amende de 5 à 5 euros</p>

Table 18: C2C stands for CamemBERT2CamemBERT. OrangeSum document 22423.

Document	<p>Le 18 octobre dernier, Jacline Mouraud se faisait connaître en publiant sur Facebook une vidéo dans laquelle elle poussait un "coup de gueule" contre le gouvernement. Aujourd'hui, la Bretonne a pris ses distances par rapport au mouvement, notamment face à d'autres figures plus radicales comme Éric Drouet. Jacline Mouraud réfléchit désormais à créer son propre parti, "la seule chose envisageable", comme elle l'explique au JDD. Nicolas Sarkozy, "le seul qui a des couilles" Cette figure des "gilets jaunes", accusée de faire le jeu de LREM estime que "le problème" d'Emmanuel Macron "c'est qu'il est jeune". "Il devrait y avoir un âge minimum pour être président : 50 ans", souligne Jacline Mouraud. Dans le JDD, elle raconte d'ailleurs avoir voté blanc lors de la dernière présidentielle. En 2007 et 2012, c'est Nicolas Sarkozy, "le seul qui a des couilles", que la figure des "gilets jaunes" avait soutenu. En attendant de se lancer, pas question pour elle en tous les cas d'être candidate aux européennes sur une liste de La République en marche.</p>
ABSTRACT	<p>Gold L'une des figures du mouvement ne sera toutefois pas candidate aux prochaines élections européennes.</p> <p>mBART Jacline Mouraud, figure des "gilets jaunes", estime que le président d'Emmanuel Macron est trop jeune pour être président.</p> <p>mBARThez Dans un entretien au JDD, la figure des "gilets jaunes" Jacline Mouraud révèle qu'elle réfléchit à créer son propre parti.</p> <p>BARThez Dans les colonnes du JDD, la figure des "gilets jaunes" explique qu'elle envisage de se présenter aux européennes sur une liste La République en marche.</p> <p>C2C Retirée de la vie politique depuis plusieurs mois, Bretonne Mouraud envisage de se lancer en politique. Et elle réfléchit à quelque chose de plus, rapporte le JDD.</p>
TITLE	<p>Gold "Gilets jaunes" : Jacline Mouraud réfléchit à créer son parti</p> <p>mBART "Gilets jaunes" : Jacline Mouraud lance son propre parti</p> <p>mBARThez "Gilets jaunes" : Jacline Mouraud prend ses distances</p> <p>BARThez La figure des "gilets jaunes" Jacline Mouraud va créer son propre parti</p> <p>C2C "Gilets jaunes" : Jacline Mouraud réfléchit à sa propre candidature</p>

Table 19: C2C stands for CamemBERT2CamemBERT. OrangeSum document 19233.

Document	<p>Invité du "Grand rendez-vous Europe 1/CNews/Les Échos dimanche 8 avril, Jean-Luc Mélenchon a appelé "à faire baisser la température dans ce pays". En cause : les menaces de mort dont il ferait l'objet, ainsi que d'autres élus LFI. Le député des Bouches-du-Rhône a confirmé avoir récemment demandé que le ministre de l'Intérieur Gérard Collomb soit entendu dans l'enquête sur un projet d'attentat d'ultra-droite où il a été cité comme cible potentielle. "Je me suis porté partie civile dans cette affaire. J'ai appris en octobre dernier qu'un groupe de gens avait l'intention de me tuer, ainsi que (le secrétaire d'État) M. Castaner". Or pendant la campagne législative de juin 2017, "j'ai demandé à être protégé" car "j'avais reçu à Marseille des menaces de mort. On me l'a refusé, et puis après je découvre que le 28 mai, ils ont arrêté ce personnage (...) Quatre mois plus tard ils en arrêtent neuf autres qui étaient toujours en action pendant ces quatre mois". "LA RECRUESCENCE D'UN EXTRÉMISME D'EXTRÊME DROITE EXTRÊMEMENT VIOLENT"Un ancien militant du groupuscule royaliste Action française en Provence, Alexandre Nisin, a été mis en examen début juillet pour association de malfaiteurs terroriste criminelle. Huit autres suspects ont été mis en examen, soupçonnés d'appartenir à son réseau. "Ni moi, ni Castaner n'avons été prévenus de rien", a déploré l'ancien candidat à la présidentielle. "Sur 17 que nous sommes au groupe La France insoumise (à l'Assemblée, ndlr), il y en a cinq qui font l'objet de menaces de mort"/(BOLD], a-t-il par ailleurs révélé. Jean-Luc Mélenchon a dénoncé "la recrudescence d'un extrémisme d'extrême droite extrêmement violent, dans toutes sortes de villes, qui va jusqu'à des tentatives d'assassinat". "L'extrême droite doit être prise au sérieux comme danger de violence et de meurtre. C'est eux qui attaquent à Montpellier un amphi d'étudiants, c'est eux qui attaquent à Tolbiac, c'est eux qui me menacent de mort. C'est eux qui font des contrôles d'identité dans la rue dans au moins deux villes. Ça suffit. Maintenant le ministre de l'Intérieur doit prendre au sérieux la menace que représentent les groupuscules radicalisés de l'extrême droite", a-t-il poursuivi."Il y a des groupes d'extrême droite qui prolifèrent dans le pays. Qui souvent ont commencé leurs premiers pas avec le Front national et qui maintenant vont au bout de cette logique", a-t-il ajouté.</p>
ABSTRACT	<p>Gold Le leader de La France insoumise (LFI) dénonce la "recrudescence" d'une "extrême droite extrêmement violent(e)" en France, qui doit être "prise au sérieux" par le gouvernement.</p> <p>mBART S'il dénonce la recrudescence d'un extrémisme d'extrême droite "extrêmement violent" dans certaines villes, le chef de file de La France insoumise (LFI), Jean-Luc Mélenchon, s'est attaqué au ministre de l'Intérieur, Gérard Collomb.</p> <p>mBARThez Au micro d'Europe 1 dimanche 8 avril, Jean-Luc Mélenchon a réagi aux menaces de mort dont il fait l'objet et dénoncé "la recrudescence d'un extrémisme d'extrême droite extrêmement violent".</p> <p>BARThez - Le chef de file de La France insoumise et ancien candidat à la présidentielle est vent debout contre le projet d'attentat déjoué à Marseille. Il estime que le ministre de l'Intérieur, Gérard Collomb, est menacé de mort par un groupe d'extrême droite.</p> <p>C2C Selon le leader de La France insoumise (LFI), le député des Bouches-du-Rhône, Jean-Luc Mélenchon, "rappelle à tous ceux suspectés d'avoir menacé d'assassiner le ministre de l'Intérieur, ce que conteste le parti et</p>
TITLE	<p>Gold VIDÉO. Cinq députés de La France insoumise font l'objet de menaces de mort, selon Jean-Luc Mélenchon</p> <p>mBART Menaces de mort : Jean-Luc Mélenchon s'en prend à Castaner</p> <p>mBARThez Jean-Luc Mélenchon dénonce les "menaces de mort" de Gérard Collomb</p> <p>BARThez VIDÉO. Jean-Luc Mélenchon dénonce les menaces de mort dont Gérard Collomb est victime</p> <p>C2C Projet VIDÉO. Menace de mort à Marseille : Jean-Luc Mélenchon menace de démissionner</p>

Table 20: C2C stands for CamemBERT2CamemBERT. OrangeSum document 22060.