ShritejShrikant_Chavan_HW6b

October 30, 2023

[1]:
```
!nvidia-smi
```

```
Mon Oct 30 01:13:01 2023
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 525.105.17   Driver Version: 525.105.17   CUDA Version: 12.0     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  Tesla T4            Off  | 00000000:00:04.0 Off |                    0 |
| N/A   33C    P8     9W /  70W |      0MiB / 15360MiB |      0%      Default |
|                               |                      |                  N/A |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
|  No running processes found                                                 |
+-----------------------------------------------------------------------------+
```

1 Translation

1.1 Outline

1. **Setting up the Environment**: Installing necessary libraries and setting up paths.
2. **Exploring and Understanding IMDB Dataset**: Understanding the structure and content of the dataset.
3. **Data Preprocessing**: Techniques to prepare the data for training, including handling different data splits and tokenization
4. **Training the Model**: Feeding data and adjusting weights.
5. **Prediction and Evaluation**: Evaluate model on test set and making predictions.

2 Setting up the Environment

[93]:
```
from pathlib import Path
if 'google.colab' in str(get_ipython()):
    from google.colab import drive
    drive.mount("/content/drive")
    !pip install datasets transformers[sentencepiece] evaluate wandb accelerate␣
 ↪-U -qq
    !pip install sacrebleu bert_score -U -qq
    base_folder = Path("/content/drive/MyDrive/NLP")

else:
    base_folder = Path("/home/harpreet/Insync/google_drive_shaannoor/data")


from transformers import AutoConfig, AutoModelForSeq2SeqLM, AutoTokenizer,␣
 ↪Seq2SeqTrainer
from transformers import AutoTokenizer, DataCollatorForSeq2Seq, pipeline
from datasets import load_dataset, DatasetDict
import evaluate
from evaluate import evaluator


import wandb
import numpy as np
import pandas as pd
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call
drive.mount("/content/drive", force_remount=True).
                            61.1/61.1 kB
1.6 MB/s eta 0:00:00
```

3 Exploring and Understanding Dataset

3.1 English_French_Translation

3.2 Load Data set

[3]:
```
kde_dataset = load_dataset('kde4', lang1='en', lang2='fr')
```

3.3 Understanding your data

[4]:
```
print(kde_dataset)
```

```
DatasetDict({
    train: Dataset({
```

```
        features: ['id', 'translation'],
        num_rows: 210173
    })
})
```

3.4 Understanding the datatype of columns

[5]:
```
kde_dataset["train"].features
```

[5]:
```
{'id': Value(dtype='string', id=None),
 'translation': Translation(languages=['en', 'fr'], id=None)}
```

3.5 Acess indivdual element

The dataset behaves like an Python array or list. We can use the familiar indexing/slicing methods to access individual elements.

[6]:
```
# get the first example of the dataset
kde_dataset["train"][42]
```

[6]:
```
{'id': '42',
 'translation': {'en': 'The Embedded Media Player',
  'fr': 'Le Lecteur de Média incorporé'}}
```

The dialogues resemble typical SMS or WhatsApp chats, complete with emojis and spaces for GIFs. The "dialogue" field presents the entire text, while the "summary" field provides the summary of the conversation.

3.6 Exploratory Data Analysis (EDA)

3.6.1 Change dataset format to Pandas

[7]:
```
# This will convert all the splits into Pandas dataframe
kde_dataset.set_format(type="pandas")
```

[8]:
```
# Get all rows from training split
df_train = kde_dataset["train"][:]
```

[9]:
```
df_train.head()
```

[9]:

	id	translation
0	0	{'en': 'Lauri Watts', 'fr': 'Lauri Watts'}
1	1	{'en': '& Lauri. Watts. mail;', 'fr': '& Lauri…
2	2	{'en': 'ROLES_OF_TRANSLATORS', 'fr': '& traduc…
3	3	{'en': '2006-02-26 3.5.1', 'fr': '2006-02-26 3…
4	4	{'en': 'The Babel & konqueror; plugin gives yo…

[10]:
```
df_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 210173 entries, 0 to 210172
Data columns (total 2 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   id           210173 non-null  object
 1   translation  210173 non-null  object
dtypes: object(2)
memory usage: 3.2+ MB
```

3.6.2 Check length of the dialogue/summary

[11]:
```
df_train['words_per_sent_en'] = df_train['translation'].apply(lambda x:␣
 ↪len(x['en'].split()))
df_train['words_per_sent_fr'] = df_train['translation'].apply(lambda x:␣
 ↪len(x['fr'].split()))
```

[12]:
```
df_train.head()
```

[12]:

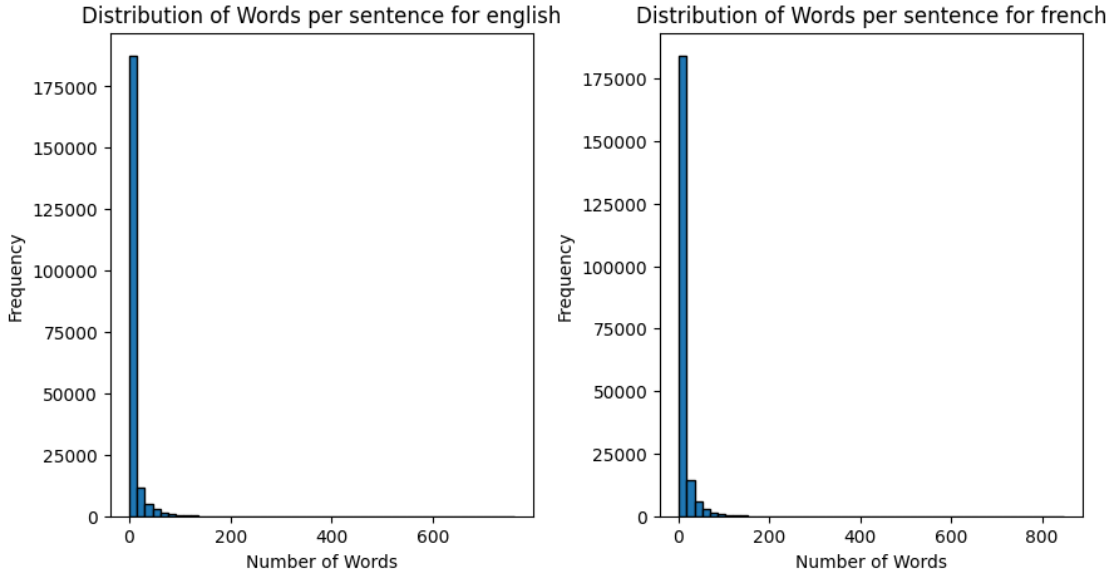
	id	translation	words_per_sent_en	words_per_sent_fr
0	0	{'en': 'Lauri Watts', 'fr': 'Lauri Watts'}	2	2
1	1	{'en': '& Lauri. Watts. mail;', 'fr': '& Lauri…	4	4
2	2	{'en': 'ROLES_OF_TRANSLATORS', 'fr': '& traduc…	1	2
3	3	{'en': '2006-02-26 3.5.1', 'fr': '2006-02-26 3…	2	2
4	4	{'en': 'The Babel & konqueror; plugin gives yo…	14	17

Plot the distribution of review length

[13]:
```
import matplotlib.pyplot as plt


def plot_two_hists(series1, series2, label1, label2):
    # Create a figure with two subplots
    fig, axs = plt.subplots(1, 2, figsize=(10, 5))

    # Plot the first histogram on the left subplot
    axs[0].hist(series1, bins=50, edgecolor="black")
    axs[0].set_xlabel("Number of Words")
    axs[0].set_ylabel("Frequency")
    axs[0].set_title(f"Distribution of Words per {label1}")
```

```
    # Plot the second histogram on the right subplot
    axs[1].hist(series2, bins=50, edgecolor="black")
    axs[1].set_xlabel("Number of Words")
    axs[1].set_ylabel("Frequency")
    axs[1].set_title(f"Distribution of Words per {label2}")

    # Adjust the spacing between subplots
    plt.subplots_adjust(wspace=0.3)

    # Display the plot
    plt.show()
```

[14]:
```
plot_two_hists(
    df_train["words_per_sent_en"], df_train["words_per_sent_fr"], "sentence for␣
 ↪english", "sentence for french"
)
```



[15]:
```
df_train["words_per_sent_en"].describe()
```

[15]:
```
count    210173.000000
mean          7.573218
std          14.894992
min           1.000000
25%           2.000000
50%           3.000000
75%           6.000000
```

```
max         759.000000
Name: words_per_sent_en, dtype: float64
```

[16]:
```
# Let us check how many sentences ahs more than 500 words

count = (df_train["words_per_sent_en"] > 500).sum()
print(f"Number of sentences with more than 400 words: {count}")
```

```
Number of sentences with more than 400 words: 5
```

[17]:
```
# Let us check how many sentences has less than 2 words

count = (df_train["words_per_sent_en"] <2).sum()
print(f"Number of sentences with less than 2 words: {count}")
```

```
Number of sentences with less than 2 words: 48165
```

[18]:
```
# Let us check how many sentences has less than 1 word - empty strings

count = (df_train["words_per_sent_en"] <1).sum()
print(f"Number of sentences with less than 2 words: {count}")
```

```
Number of sentences with less than 2 words: 0
```

[19]:
```
df_train[df_train["words_per_sent_fr"]<2]
```

[19]:

	id	translation	words_per_sent_en	words_per_sent_fr
5	5	{'en': 'KDE', 'fr': 'KDE'}	1	1
6	6	{'en': 'kdeaddons', 'fr': 'kdeaddons'}	1	1
7	7	{'en': 'konqueror', 'fr': 'konqueror'}	1	1
9	9	{'en': 'babelfish', 'fr': 'babelfish'}	1	1
10	10	{'en': 'translate', 'fr': 'traduction'}	1	1
…	…	…	…	…
210137	210137	{'en': 'Games', 'fr': 'JeuxPhonon::'}	1	1
210138	210138	{'en': 'Accessibility', 'fr': 'Accessibilité'}		
210156	210156	{'en': 'aRts', 'fr': 'aRts'}		
210161	210161	{'en': 'Volume', 'fr': 'Volume'}		
210172	210172	{'en': 'PulseAudio', 'fr': 'PulseAudio'}		

```
210138                  1                   1
210156                  1                   1
210161                  1                   1
210172                  1                   1

[31404 rows x 4 columns]
```

3.6.3 Reset dataset format

[20]:
```
kde_dataset.reset_format()
```

[21]:
```
kde_dataset
```

[21]:
```
DatasetDict({
    train: Dataset({
        features: ['id', 'translation'],
        num_rows: 210173
    })
})
```

4 Data Pre-processing

4.1 Create train, valid, test splits

[22]:
```
test_val_splits = kde_dataset['train'].train_test_split(test_size=0.4, seed=42)
train_split= test_val_splits['train']
test_val_splits = test_val_splits['test'].train_test_split(test_size=0.5,␣
 ↪seed=42,)
val_split = test_val_splits['train']
test_split = test_val_splits['test']
```

4.2 Create small subset for experimentation

[23]:
```
train_split_small = train_split.shuffle(seed=42).select(range(1000))
val_split_small = val_split.shuffle(seed=42).select(range(500))
test_split_small = test_split.shuffle(seed=42).select(range(500))
```

[24]:
```
# combine train, val splits into one dataset
train_val_subset = DatasetDict({'train': train_split_small, 'val':␣
 ↪val_split_small})

# create test dataset from test split
test_subset= DatasetDict({'test': test_split_small})
```

[25]:
```
train_val_subset
```

[25]:
```
DatasetDict({
    train: Dataset({
        features: ['id', 'translation'],
        num_rows: 1000
    })
    val: Dataset({
        features: ['id', 'translation'],
        num_rows: 500
    })
})
```

[26]:
```
test_subset
```

[26]:
```
DatasetDict({
    test: Dataset({
        features: ['id', 'translation'],
        num_rows: 500
    })
})
```

[27]:
```
train_val_subset["train"]["translation"][1]
```

[27]:
```
{'en': 'Upper Window: View on Printers, both Real and Virtual',
 'fr': "Fenêtre supérieure & #160;: Vue de l'ensemble des imprimantes, réelles
et virtuelles"}
```

4.3 Tokenization

4.3.1 Load pre-trained Tokenizer

[28]:
```
checkpoint = 'Helsinki-NLP/opus-mt-en-fr'
tokenizer = AutoTokenizer.from_pretrained(checkpoint)
```

```
/usr/local/lib/python3.10/dist-
packages/transformers/models/marian/tokenization_marian.py:197: UserWarning:
Recommended: pip install sacremoses.
  warnings.warn("Recommended: pip install sacremoses.")
```

[29]:
```
# tokenizer??
```

4.3.2 Understanding tokenizer

[30]:
```
text = [
    "Upper Window: View on Printers, both Real and Virtual",
    "Fenêtre supérieure & #160;: Vue de l'ensemble des imprimantes, réelles et␣
 ↪virtuelles",
]
```

[31]:
```
# get the vocab size
print(f"Pretrained tokenizer vocab size {tokenizer.vocab_size}")
```

```
Pretrained tokenizer vocab size 59514
```

[32]:
```
encoded_text = tokenizer(
    text, padding=True, truncation=True, return_tensors="pt")
```

[33]:
```
encoded_text
```

[33]:
```
{'input_ids': tensor([[21364, 38007,    37,  3689,    30, 36712,     9,     2,
533,  7770,
            10, 21821,     0, 59513, 59513, 59513, 59513, 59513, 59513, 59513,
         59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513,
         59513, 59513],
        [37412,   501,    62,   746, 32076,  1935,   402, 30362,  2069,    50,
            37,   461,  3773,     5,    14,     6,  8409,   122,  4995,    13,
          3497, 11585,  3172,     9,     2,   906,  1295,    11, 13747,   477,
           114,     0]]), 'attention_mask': tensor([[1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
         0, 0, 0, 0, 0, 0, 0, 0],
        [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
         1, 1, 1, 1, 1, 1, 1, 1]])}
```

[34]:
```
tokens_first_sentence = tokenizer.convert_ids_to_tokens(
    encoded_text.input_ids[0])
tokens_second_sentence = tokenizer.convert_ids_to_tokens(
    encoded_text.input_ids[1])

print(tokens_first_sentence)
print(tokens_second_sentence)
```

```
[' Upper', ' Window', ':', ' View', ' on', ' Printer', 's', ',', ' both',
' Real', ' and', ' Virtual', '</s>', '<pad>', '<pad>', '<pad>', '<pad>',
'<pad>', '<pad>', '<pad>', '<pad>', '<pad>', '<pad>', '<pad>', '<pad>', '<pad>',
'<pad>', '<pad>', '<pad>', '<pad>', '<pad>']
[' Fen', 'être', ' s', 'up', 'érie', 'ure', ' &', ' #1', '60', ';', ':', ' V',
'ue', ' de', ' l', "'", 'ense', 'm', 'ble', ' des', ' imp', 'rim', 'ante', 's',
',', ' ré', 'elles', ' et', ' virtue', 'll', 'es', '</s>']
```

[35]:
```
tokenizer.convert_tokens_to_string(tokens_first_sentence)
```

[35]:
```
'Upper Window: View on Printers, both Real and Virtual</s> <pad> <pad> <pad>
<pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad>
<pad> <pad> <pad>'
```

[36]:
```
tokenizer.convert_tokens_to_string(tokens_second_sentence)
```

[36]:
```
"Fenêtre supérieure & #160;: Vue de l'ensemble des imprimantes, réelles et
virtuelles</s>"
```

[37]:
```
special_tokens = tokenizer.all_special_tokens
special_tokens_ids = tokenizer.all_special_ids
```

[38]:
```
import pandas as pd
```

[39]:
```
df = pd.DataFrame(
    {"special_tokens": special_tokens, "special_tokens_ids": special_tokens_ids}
)
```

[40]:
```
pd.set_option("display.max_rows", None)
```

[41]:
```
df
```

[41]:

	special_tokens	special_tokens_ids
0	</s>	0
1	<unk>	1
2	<pad>	59513

4.3.3 Create function for Tokenizer

[43]:
```
# we do not need to add padding
# padding and conversion to Pytorch Tensor is handled by the data collators
# in more efficient manner

max_length = 128
def tokenize_fn(batch):
    inputs = [example['en'] for example in batch['translation']]
    targets = [example['fr'] for example in batch['translation']]
    model_inputs = tokenizer(text = inputs, text_target=targets, truncation =␣
 ↪True, max_length=max_length)
    return model_inputs
    # CODE HERE
```

4.3.4 Use map function to apply tokenization to all splits

[44]:
```
tokenized_dataset = train_val_subset.map(tokenize_fn,
                                batched = True,
                                remove_columns=train_val_subset['train'].
 ↪column_names)
```

```
Map:   0%|          | 0/1000 [00:00<?, ? examples/s]
Map:   0%|          | 0/500 [00:00<?, ? examples/s]
```

[45]:
```
tokenized_dataset.set_format(type="torch")
```

[46]:
```
tokenized_dataset
```

[46]:
```
DatasetDict({
    train: Dataset({
        features: ['input_ids', 'attention_mask', 'labels'],
        num_rows: 1000
    })
    val: Dataset({
        features: ['input_ids', 'attention_mask', 'labels'],
        num_rows: 500
    })
})
```

[47]:
```
tokenized_dataset["train"].features
```

[47]:
```
{'input_ids': Sequence(feature=Value(dtype='int32', id=None), length=-1,
id=None),
 'attention_mask': Sequence(feature=Value(dtype='int8', id=None), length=-1,
id=None),
 'labels': Sequence(feature=Value(dtype='int64', id=None), length=-1, id=None)}
```

[48]:
```
print(len(tokenized_dataset["train"]["input_ids"][2]))
print(len(tokenized_dataset["train"]["input_ids"][1]))
```

```
2
13
```

[49]:
```
print(
    tokenizer.convert_ids_to_tokens(
        tokenized_dataset["train"]["input_ids"][0][0:10]
    )
)
```

```
[' I', 'dent', 'ifier', ':', '</s>']
```

[50]:
```
print(
    tokenizer.convert_ids_to_tokens(
        tokenized_dataset["train"]["labels"][0][0:10])
)
```

```
[' Identifiant', ' &', ' #160;:', '</s>']
```

5 Model Training

5.1 Model Config File

[51]:
```
config = AutoConfig.from_pretrained(checkpoint)
```

[52]:
```
config
```

[52]:
```
MarianConfig {
  "_name_or_path": "Helsinki-NLP/opus-mt-en-fr",
  "_num_labels": 3,
  "activation_dropout": 0.0,
  "activation_function": "swish",
  "add_bias_logits": false,
  "add_final_layer_norm": false,
  "architectures": [
    "MarianMTModel"
  ],
  "attention_dropout": 0.0,
  "bad_words_ids": [
    [
      59513
    ]
  ],
  "bos_token_id": 0,
  "classif_dropout": 0.0,
  "classifier_dropout": 0.0,
  "d_model": 512,
  "decoder_attention_heads": 8,
  "decoder_ffn_dim": 2048,
  "decoder_layerdrop": 0.0,
  "decoder_layers": 6,
  "decoder_start_token_id": 59513,
  "decoder_vocab_size": 59514,
  "dropout": 0.1,
  "encoder_attention_heads": 8,
  "encoder_ffn_dim": 2048,
  "encoder_layerdrop": 0.0,
  "encoder_layers": 6,
  "eos_token_id": 0,
  "forced_eos_token_id": 0,
  "gradient_checkpointing": false,
  "id2label": {
    "0": "LABEL_0",
    "1": "LABEL_1",
    "2": "LABEL_2"
  },
```

```
  "init_std": 0.02,
  "is_encoder_decoder": true,
  "label2id": {
    "LABEL_0": 0,
    "LABEL_1": 1,
    "LABEL_2": 2
  },
  "max_length": 512,
  "max_position_embeddings": 512,
  "model_type": "marian",
  "normalize_before": false,
  "normalize_embedding": false,
  "num_beams": 4,
  "num_hidden_layers": 6,
  "pad_token_id": 59513,
  "scale_embedding": true,
  "share_encoder_decoder_embeddings": true,
  "static_position_embeddings": true,
  "transformers_version": "4.34.1",
  "use_cache": true,
  "vocab_size": 59514
}
```

[53]:
```
from transformers import GenerationConfig

# generation_config = GenerationConfig.from_model_config(config)
generation_config = GenerationConfig.from_pretrained(checkpoint)
```

```
Downloading (…)neration_config.json:   0%|          | 0.00/293 [00:00<?, ?B/s]
```

[54]:
```
generation_config
```

[54]:
```
GenerationConfig {
  "bad_words_ids": [
    [
      59513
    ]
  ],
  "bos_token_id": 0,
  "decoder_start_token_id": 59513,
  "eos_token_id": 0,
  "forced_eos_token_id": 0,
  "max_length": 512,
  "num_beams": 4,
  "pad_token_id": 59513,
  "renormalize_logits": true
}
```

5.2 Download pre-trained model

[55]:
```
model = AutoModelForSeq2SeqLM.from_pretrained(
    checkpoint,
    config=config,
)
```

```
Downloading pytorch_model.bin:   0%|          | 0.00/301M [00:00<?, ?B/s]
```

5.3 Model Input/Collate Function

[56]:
```
data_collator = DataCollatorForSeq2Seq(
    tokenizer=tokenizer,
    model=model,
)
```

[57]:
```
train_val_subset['train'][100]
```

[57]:
```
{'id': '34366',
 'translation': {'en': 'Basic Command Set',
  'fr': 'Ensemble de commandes basiques'}}
```

[58]:
```
features = [tokenized_dataset["train"][i] for i in range(3)]
features
```

[58]:
```
[{'input_ids': tensor([  47, 9314, 8245,   37,    0]),
  'attention_mask': tensor([1, 1, 1, 1, 1]),
  'labels': tensor([41612,   402, 38492,     0])},
 {'input_ids': tensor([21364, 38007,    37,  3689,    30, 36712,     9,     2,
533,  7770,
            10, 21821,     0]),
  'attention_mask': tensor([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]),
  'labels': tensor([  304,   162,   501,  4517,   402, 38492, 14781,     5,
14,     6,
           1420,    13, 29719,     9,     2, 12782,    11, 18769,     9,
0])},
 {'input_ids': tensor([41704,     0]),
  'attention_mask': tensor([1, 1]),
  'labels': tensor([12323,     0])}]
```

[59]:
```
model_input = data_collator(features)
model_input.keys()
```

[59]:
```
dict_keys(['input_ids', 'attention_mask', 'labels', 'decoder_input_ids'])
```

- Notice that data_collator has added the decoder_input_ids

[ ]:
```

```

[60]:
```
print(model_input)
```

```
{'input_ids': tensor([[   47,  9314,  8245,    37,     0, 59513, 59513, 59513,
59513, 59513,
         59513, 59513, 59513],
        [21364, 38007,    37,  3689,    30, 36712,     9,     2,   533,  7770,
            10, 21821,     0],
        [41704,     0, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513,
         59513, 59513, 59513]]), 'attention_mask': tensor([[1, 1, 1, 1, 1, 0, 0,
0, 0, 0, 0, 0, 0],
        [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1],
        [1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]]), 'labels': tensor([[41612,
402, 38492,     0,  -100,  -100,  -100,  -100,  -100,  -100,
          -100,  -100,  -100,  -100,  -100,  -100,  -100,  -100,  -100,  -100],
        [  304,   162,   501,  4517,   402, 38492, 14781,     5,    14,     6,
          1420,    13, 29719,     9,     2, 12782,    11, 18769,     9,     0],
        [12323,     0,  -100,  -100,  -100,  -100,  -100,  -100,  -100,  -100,
          -100,  -100,  -100,  -100,  -100,  -100,  -100,  -100,  -100,
-100]]), 'decoder_input_ids': tensor([[59513, 41612,   402, 38492,     0, 59513,
59513, 59513, 59513, 59513,
         59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513],
        [59513,   304,   162,   501,  4517,   402, 38492, 14781,     5,    14,
             6,  1420,    13, 29719,     9,     2, 12782,    11, 18769,     9],
        [59513, 12323,     0, 59513, 59513, 59513, 59513, 59513, 59513, 59513,
         59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513,
59513]])}
```

[ ]:
```
print(model_input)
```

```
{'input_ids': tensor([[   47,  9314,  8245,    37,     0, 59513, 59513, 59513,
59513, 59513,
         59513, 59513, 59513],
        [21364, 38007,    37,  3689,    30, 36712,     9,     2,   533,  7770,
            10, 21821,     0],
        [41704,     0, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513,
         59513, 59513, 59513]]), 'attention_mask': tensor([[1, 1, 1, 1, 1, 0, 0,
0, 0, 0, 0, 0, 0],
        [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1],
        [1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]]), 'labels': tensor([[41612,
402, 38492,     0,  -100,  -100,  -100,  -100,  -100,  -100,
          -100,  -100,  -100,  -100,  -100,  -100,  -100,  -100,  -100,  -100],
        [  304,   162,   501,  4517,   402, 38492, 14781,     5,    14,     6,
          1420,    13, 29719,     9,     2, 12782,    11, 18769,     9,     0],
        [12323,     0,  -100,  -100,  -100,  -100,  -100,  -100,  -100,  -100,
          -100,  -100,  -100,  -100,  -100,  -100,  -100,  -100,  -100,
-100]]), 'decoder_input_ids': tensor([[59513, 41612,   402, 38492,     0, 59513,
59513, 59513, 59513, 59513,
         59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513],
```

```
        [59513,   304,   162,   501,  4517,   402, 38492, 14781,     5,    14,
             6,  1420,    13, 29719,     9,     2, 12782,    11, 18769,     9],
        [59513, 12323,     0, 59513, 59513, 59513, 59513, 59513, 59513, 59513,
         59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513,
59513]])}
```

- **Note that 59513 is both the start token and pad token for the decoder input ids**
- **Also note that decoder input ids are shifted version of the labels**

[61]:
```
print(model_input.decoder_input_ids[0][0:10])
print(model_input.labels[0][0:10])
```

```
tensor([59513, 41612,   402, 38492,     0, 59513, 59513, 59513, 59513, 59513])
tensor([41612,   402, 38492,     0,  -100,  -100,  -100,  -100,  -100,  -100])
```

[62]:
```
print(model_input.decoder_input_ids[0][-20:])
print(model_input.labels[0][-20:])
```

```
tensor([59513, 41612,   402, 38492,     0, 59513, 59513, 59513, 59513, 59513,
        59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513])
tensor([41612,   402, 38492,     0,  -100,  -100,  -100,  -100,  -100,  -100,
         -100,  -100,  -100,  -100,  -100,  -100,  -100,  -100,  -100,  -100])
```

[63]:
```
print(model_input.decoder_input_ids[1][0:10])
print(model_input.labels[1][0:10])
```

```
tensor([59513,   304,   162,   501,  4517,   402, 38492, 14781,     5,    14])
tensor([  304,   162,   501,  4517,   402, 38492, 14781,     5,    14,     6])
```

[64]:
```
print(model_input.decoder_input_ids[1][-20:])
print(model_input.labels[1][-20:])
```

```
tensor([59513,   304,   162,   501,  4517,   402, 38492, 14781,     5,    14,
            6,  1420,    13, 29719,     9,     2, 12782,    11, 18769,     9])
tensor([  304,   162,   501,  4517,   402, 38492, 14781,     5,    14,     6,
         1420,    13, 29719,     9,     2, 12782,    11, 18769,     9,     0])
```

[65]:
```
print(tokenizer.convert_ids_to_tokens(model_input.decoder_input_ids[0][0:10]))
print(tokenizer.convert_ids_to_tokens(model_input.labels[0][0:10]))
```

```
['<pad>', ' Identifiant', ' &', ' #160;:', '</s>', '<pad>', '<pad>', '<pad>',
'<pad>', '<pad>']
[' Identifiant', ' &', ' #160;:', '</s>', '<unk>', '<unk>', '<unk>', '<unk>',
'<unk>', '<unk>']
```

[66]:
```
print(tokenizer.convert_ids_to_tokens(model_input.decoder_input_ids[0][-20:]))
try:
    print(tokenizer.convert_ids_to_tokens(model_input.labels[0][-20:]))
```

```
except:
    print(f"No tokens corresponding to ids")
```

```
['<pad>', ' Identifiant', ' &', ' #160;:', '</s>', '<pad>', '<pad>', '<pad>',
'<pad>', '<pad>', '<pad>', '<pad>', '<pad>', '<pad>', '<pad>', '<pad>', '<pad>',
'<pad>', '<pad>', '<pad>']
[' Identifiant', ' &', ' #160;:', '</s>', '<unk>', '<unk>', '<unk>', '<unk>',
'<unk>', '<unk>', '<unk>', '<unk>', '<unk>', '<unk>', '<unk>', '<unk>', '<unk>',
'<unk>', '<unk>', '<unk>']
```

5.4 Understanding Model Output

[67]:
```
# model output
model_output = model(**model_input)
```

[68]:
```
# keys in model output
model_output.keys()
```

[68]:
```
odict_keys(['loss', 'logits', 'encoder_last_hidden_state'])
```

[69]:
```
# let us look at logits
model_output.logits.shape
```

[69]:
```
torch.Size([3, 20, 59514])
```

[70]:
```
logits = model_output.logits.detach().cpu().numpy()
```

[71]:
```
# Greedy decoding
preds = np.argmax(logits, axis = -1)
```

[72]:
```
preds.shape
```

[72]:
```
(3, 20)
```

[73]:
```
preds
```

[73]:
```
array([[16235,    37, 38492,     0,     0,     0,     0,     0,     0,
           47,    47,    47,    47,    47,    47,    47,    47,    47,
           47,    47],
       [  304,   162,   501,  4517,    78, 38492,   498,    36,    14,
            6, 41599,    13, 29719,     9,     2, 12782,    11, 18769,
            9,     0],
       [12323,     0,     0,     0,     0,     0, 12323, 12323, 12323,
        12323, 12323, 12323, 12323, 12323, 12323, 12323, 12323, 12323,
        12323, 12323]])
```

[74]:
```
tokenizer.convert_tokens_to_string(tokenizer.convert_ids_to_tokens(preds[0]))
```

[74]:
```
'Identification: #160;:</s> </s> </s> </s> </s> </s> I I I I I I I I I I I'
```

[75]:
```
def get_label_strings(labels):
    return tokenizer.convert_tokens_to_string(tokenizer.
 ↪convert_ids_to_tokens(labels))
```

[76]:
```
labels = [get_label_strings(label) for label in model_input['labels']]
```

[77]:
```
labels
```

[77]:
```
['Identifiant & #160;:</s> <unk> <unk> <unk> <unk> <unk> <unk> <unk> <unk> <unk>
<unk> <unk> <unk> <unk> <unk> <unk> <unk> <unk>',
 "Fenêtre supérieure & #160;: Vue de l'ensemble des imprimantes, réelles et
virtuelles</s>",
 'Aperçu</s> <unk> <unk> <unk> <unk> <unk> <unk> <unk> <unk> <unk> <unk> <unk>
<unk> <unk> <unk> <unk> <unk> <unk> <unk> <unk>']
```

[78]:
```
preds = [ get_label_strings(pred) for pred in preds]
preds
```

[78]:
```
['Identification: #160;:</s> </s> </s> </s> </s> </s> I I I I I I I I I I I',
 "Fenêtre supérieure : #160;: vue sur l'imprimante des imprimantes, réelles et
virtuelles</s>",
 'Aperçu</s> </s> </s> </s> </s> Aperçu Aperçu Aperçu Aperçu Aperçu Aperçu
Aperçu Aperçu Aperçu Aperçu Aperçu Aperçu Aperçu Aperçu']
```

[79]:
```
final_string_preds=[]
for example in preds:
    string_preds =[]
    for s in example.split():
        string_preds.append(s)
        if '</s>' in s:
            break
    final_string_preds.append(' '.join(string_preds))

final_string_preds
```

[79]:
```
['Identification: #160;:</s>',
 "Fenêtre supérieure : #160;: vue sur l'imprimante des imprimantes, réelles et
virtuelles</s>",
 'Aperçu</s>']
```

[80]:
```
# generate prediction from logits using beam search
generated_predictions = model.generate(
    **model_input, generation_config=generation_config, max_length=128
)
```

[81]:
```
generated_predictions.shape
```

[81]:
```
torch.Size([3, 25])
```

[82]:
```
generated_predictions
```

[82]:
```
tensor([[59513, 41612,   402, 38492,     0, 59513, 59513, 59513, 59513, 59513,
         59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513,
            47,  9314, 18458,    37,     0],
        [59513,   304,   162,   501,  4517,   402, 38492, 14781,     5,    14,
             6,  1420,    13, 29719,     9,     2, 12782,    11, 18769,     9,
             0, 59513, 59513, 59513, 59513],
        [59513, 12323,     0, 59513, 59513, 59513, 59513, 59513, 59513, 59513,
         59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513, 59513,
         12323,     0, 59513, 59513, 59513]])
```

[83]:
```
generated_predictions.shape
```

[83]:
```
torch.Size([3, 25])
```

[84]:
```
tokens = tokenizer.convert_ids_to_tokens(generated_predictions[0])
print(tokens)
```

```
['<pad>', ' Identifiant', ' &', ' #160;:', '</s>', '<pad>', '<pad>', '<pad>',
'<pad>', '<pad>', '<pad>', '<pad>', '<pad>', '<pad>', '<pad>', '<pad>', '<pad>',
'<pad>', '<pad>', '<pad>', ' I', 'dent', 'ificateur', ':', '</s>']
```

[85]:
```
translation = tokenizer.convert_tokens_to_string(tokens)
print(translation)
```

```
<pad> Identifiant & #160;:</s> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad>
<pad> <pad> <pad> <pad> <pad> <pad> <pad> Identificateur:</s>
```

[86]:
```
tokenizer.decode(generated_predictions[1])
```

[86]:
```
"<pad> Fenêtre supérieure & #160;: Vue de l'ensemble des imprimantes, réelles et
virtuelles</s> <pad> <pad> <pad> <pad>"
```

[87]:
```
tokenizer.decode(generated_predictions[1], skip_special_tokens=True)
```

[87]:
```
"Fenêtre supérieure & #160;: Vue de l'ensemble des imprimantes, réelles et
virtuelles"
```

[88]:
```
# decode tokens to generate translation
tokenizer.batch_decode(generated_predictions)
```

[88]:
```
['<pad> Identifiant & #160;:</s> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad>
<pad> <pad> <pad> <pad> <pad> <pad> <pad> Identificateur:</s>',
 "<pad> Fenêtre supérieure & #160;: Vue de l'ensemble des imprimantes, réelles
et virtuelles</s> <pad> <pad> <pad> <pad>",
 '<pad> Aperçu</s> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad>
<pad> <pad> <pad> <pad> <pad> <pad> <pad> Aperçu</s> <pad> <pad> <pad>']
```

[89]:
```
# decode tokens to generate summary
tokenizer.batch_decode(generated_predictions, skip_special_tokens=True)
```

[89]:
```
['Identifiant & #160;: Identificateur:',
 "Fenêtre supérieure & #160;: Vue de l'ensemble des imprimantes, réelles et
virtuelles",
 'Aperçu Aperçu']
```

5.5 Evaluation metric(s)

5.5.1 Function to compute metric

[94]:
```
bleu_metric = evaluate.load("sacrebleu")
bert_metric = evaluate.load('bertscore')

def compute_metrics(preds_and_labels):
    # preds are not logits but token ids
    # api is inconsistent here
    # we are not simply using argmax bu use 'beam search'
    preds, labels = preds_and_labels

    # convert predictions into words
    decoded_preds = tokenizer.batch_decode(preds, skip_special_tokens=True)

    # for any -100 label, replace with pad token id
    labels = np.where( labels != -100, labels, tokenizer.pad_token_id )

    # convert labels into words
    decoded_labels = tokenizer.batch_decode(labels, skip_special_tokens= True)

    # get rid of extra whitespace
    # and also, put targets into lists

    decoded_preds_cleaned = [pred.strip() for pred in decoded_preds]
    decoded_labels_cleaned = [label.strip() for label in decoded_labels]

    bleu_score = bleu_metric.compute(predictions=decoded_preds_cleaned,␣
 ↪references=decoded_labels_cleaned)
    bert_score = bert_metric.compute(predictions=decoded_preds_cleaned,␣
 ↪references=decoded_labels_cleaned, lang='fr')
```

```
    return{'bleu_score:': bleu_score['score'], 'bert_score': np.
↪mean(bert_score['f1'])}
    # CODE HERE
```

5.6 Set up Logger for experiments

[95]:
```
wandb.login()
# Set project name for logging
%env WANDB_PROJECT = nlp_course_fall_2023-translation
```

```
<IPython.core.display.Javascript object>
```

```
wandb: Appending key for api.wandb.ai to your netrc file:
/root/.netrc
```

```
env: WANDB_PROJECT=nlp_course_fall_2023-translation
```

5.7 Hyperparameters and Checkpointing

[96]:
```
from transformers import Seq2SeqTrainingArguments

# Define the directory where model checkpoints will be saved
model_folder = base_folder / "models" / "nlp_spring_2023/kde4/opus-mt-en-fr"

# Create the directory if it doesn't exist
model_folder.mkdir(exist_ok=True, parents=True)

# Configure training parameters
training_args = Seq2SeqTrainingArguments(
    # Training-specific configurations
    num_train_epochs=1,  # Total number of training epochs
    weight_decay=0.01,  # Apply L2 regularization to prevent overfitting
    learning_rate=5e-5,  # Step size for the optimizer during training
    optim="adamw_torch",  # Optimizer,
    warmup_steps=10,
    predict_with_generate=True,
    generation_config=generation_config,
    # memory and speed related arguments
    # Number of samples per training batch for each device
    per_device_train_batch_size=16,
    per_device_eval_batch_size=16,  # Number of samples per eval batch for each␣
↪device

    gradient_checkpointing=True,  # memory
    # fp16 = True, # Speed
    # bf16=True,
    # tf32=True, # speed
```

```
    # evaluation settings
    output_dir=str(model_folder),  # Directory to save model checkpoints
    evaluation_strategy="steps",  # Evaluate model at specified step intervals
    eval_steps=10,  # Perform evaluation every 10 training steps
    # Checkpoint settings
    save_strategy="steps",  # Save model checkpoint at specified step intervals
    save_steps=10,  # Save a model checkpoint every 10 training steps
    load_best_model_at_end=True,  # Reload the best model at the end of training
    save_total_limit=2,  # Retain only the best and the most recent model␣
 ↪checkpoints
    # metric_for_best_model=,
    # greater_is_better=,
    # Experiment logging configurations (commented out in this example)
    logging_strategy="steps",
    logging_steps=10,
    report_to="wandb",  # Log metrics and results to Weights & Biases platform
    # Experiment name for Weights & Biases
    run_name="translation-exp1",
)
```

5.8 Initialize Trainer

[97]:
```
# initialize trainer
trainer = Seq2SeqTrainer(
    model=model,
    args=training_args,
    train_dataset=tokenized_dataset["train"],
    eval_dataset=tokenized_dataset["val"],
    compute_metrics=compute_metrics,
    tokenizer=tokenizer,
    data_collator=data_collator,
)
```

5.9 Start Training

[98]:
```
import gc
import torch
torch.cuda.empty_cache()
gc.collect()
```

[98]: 1301

[99]:
```
trainer.train()  # start training
```

<IPython.core.display.HTML object>

```
wandb: Currently logged in as: shritej24c
(redeem_team). Use `wandb login --relogin` to force relogin

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(

<IPython.core.display.HTML object>

Downloading (…)okenizer_config.json:   0%|          | 0.00/29.0 [00:00<?, ?B/s]

Downloading (…)lve/main/config.json:   0%|          | 0.00/625 [00:00<?, ?B/s]

Downloading (…)solve/main/vocab.txt:   0%|          | 0.00/996k [00:00<?, ?B/s]

Downloading (…)/main/tokenizer.json:   0%|          | 0.00/1.96M [00:00<?, ?B/s]

Downloading model.safetensors:   0%|          | 0.00/714M [00:00<?, ?B/s]

/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
```

```
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
```

[99]:
```
TrainOutput(global_step=63, training_loss=1.6225580639309354,
metrics={'train_runtime': 517.4676, 'train_samples_per_second': 1.932,
'train_steps_per_second': 0.122, 'total_flos': 15072060506112.0, 'train_loss':
1.6225580639309354, 'epoch': 1.0})
```

5.10 Evaluation

5.10.1 Check performance on validation set

[100]:
```
trainer.evaluate(tokenized_dataset["val"])
```

```
<IPython.core.display.HTML object>
```

[100]:
```
{'eval_loss': 1.5204508304595947,
 'eval_bleu_score:': 41.166549338334434,
 'eval_bert_score': 0.8666297186017037,
 'eval_runtime': 81.766,
 'eval_samples_per_second': 6.115,
 'eval_steps_per_second': 0.391,
 'epoch': 1.0}
```

[ ]:
```
trainer.evaluate(tokenized_dataset["val"])
```

```
  0%|          | 0/32 [00:00<?, ?it/s]
```

[ ]:
```
{'eval_loss': 1.5268305540084839,
 'eval_bleu_score:': 41.43463120346539,
 'eval_bert_score': 0.8658450670838356,
 'eval_runtime': 55.277,
 'eval_samples_per_second': 9.045,
 'eval_steps_per_second': 0.579,
 'epoch': 1.0}
```

[102]:
```
wandb.finish()  # stop logging
```

```
<IPython.core.display.HTML object>
```

```
VBox(children=(Label(value='0.002 MB of 0.002 MB uploaded (0.000 MB␣
↪deduped)\r'), FloatProgress(value=1.0, max…
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

[ ]:
```
wandb.finish()  # stop logging
```

```
<IPython.core.display.HTML object>
```

```
VBox(children=(Label(value='0.002 MB of 0.002 MB uploaded (0.000 MB␣
↪deduped)\r'), FloatProgress(value=1.0, max…
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

5.10.2 Check the best saved model

[103]:
```
# After training, let us check the best checkpoint
# We need this for Predioctions and Evaluations
best_model_checkpoint_step = trainer.state.best_model_checkpoint.split("-")[-1]
print(f"The best model was saved at step {best_model_checkpoint_step}.")
```

```
The best model was saved at step 60.
```

6 Inference

6.1 Test Set Evaluation

[108]:
```
model_folder = base_folder / "models" / "nlp_spring_2023/kde4/opus-mt-en-fr"
checkpoint = str(model_folder / "checkpoint-60")
test_subset
```

[108]:
```
DatasetDict({
    test: Dataset({
        features: ['id', 'translation'],
        num_rows: 500
    })
})
```

[109]:
```
test_data_flattened = test_subset["test"].map(lambda example: {'en':␣
↪example['translation']['en'], 'fr': example['translation']['fr']})
```

[110]:
```
test_data_flattened
```

[110]:
```
Dataset({
    features: ['id', 'translation', 'en', 'fr'],
    num_rows: 500
})
```

[111]:
```
task_evaluator = evaluator("translation")
gen_kwargs = {"length_penalty": 0.8, "num_beams": 8, "max_length": 128}
eval_results = task_evaluator.compute(
    model_or_pipeline=checkpoint,
    tokenizer=checkpoint,
    data=test_data_flattened,
    input_column='en',
    label_column='fr',
    generation_kwargs=gen_kwargs,
    device=0,
)
```

```
/usr/local/lib/python3.10/dist-
packages/transformers/models/marian/tokenization_marian.py:197: UserWarning:
Recommended: pip install sacremoses.
  warnings.warn("Recommended: pip install sacremoses.")
Downloading builder script:   0%|          | 0.00/5.94k [00:00<?, ?B/s]
Downloading extra modules:   0%|          | 0.00/1.55k [00:00<?, ?B/s]
Downloading extra modules:   0%|          | 0.00/3.34k [00:00<?, ?B/s]
Your input_length: 130 is bigger than 0.9 * max_length: 128. You might consider
increasing your max_length manually, e.g. translator('…', max_length=400)
```

[112]:
```
eval_results
```

[112]:
```
{'bleu': 0.42696869676160426,
 'precisions': [0.7095400340715503,
  0.5355100095328885,
  0.4212889593238246,
```

```
 0.32914719626168226],
'brevity_penalty': 0.8911839411730454,
'length_ratio': 0.8966965820126026,
'translation_length': 4696,
'reference_length': 5237,
'total_time_in_seconds': 110.03906021500006,
'samples_per_second': 4.543841059920667,
'latency_in_seconds': 0.22007812043000013}
```

[ ]:
```
eval_results
```

[ ]:
```
{'bleu': 0.4297887382432084,
 'precisions': [0.6963209311993348,
  0.5226165622825331,
  0.4125192406362237,
  0.32389580973952437],
 'brevity_penalty': 0.9152600196375821,
 'length_ratio': 0.9186557189230475,
 'translation_length': 4811,
 'reference_length': 5237,
 'total_time_in_seconds': 73.25119658300537,
 'samples_per_second': 6.825827062543882,
 'latency_in_seconds': 0.14650239316601074}
```

6.2 Prediction for individual or small list of examples

6.3 Create pipelne for inference

[113]:
```
custom_pipeline = pipeline(
    task="translation",
    model=checkpoint,
    tokenizer=checkpoint,
    device='cpu',
    framework = 'pt'
)
```

```
/usr/local/lib/python3.10/dist-
packages/transformers/models/marian/tokenization_marian.py:197: UserWarning:
Recommended: pip install sacremoses.
  warnings.warn("Recommended: pip install sacremoses.")
```

[114]:
```
sample = test_data_flattened["en"][420]
sample
```

[114]:
```
'Find matching recipes'
```

[115]:
```
translation = custom_pipeline(sample)
translation
```

[115]:
```
[{'translation_text': 'Trouver des recettes correspondantes'}]
```

[116]:
```
samples = test_data_flattened["en"][420:423]
samples
```

[116]:
```
['Find matching recipes',
 '%1x%2x%3 brick, %4 shuffling moves',
 'The following sections will describe, in detail, each element. Everything from
where they can go to what goes in them is layed out in an easy-to-follow
manner.']
```

[117]:
```
gen_kwargs = {"length_penalty": 0.8, "num_beams": 8, "max_length": 128}
translations = custom_pipeline(samples, **gen_kwargs)
translations
```

[117]:
```
[{'translation_text': 'Trouver des recettes correspondantes'},
 {'translation_text': '%1x%2x%3 brique, %4 mouvements de shuffling'},
 {'translation_text': "Les sections suivantes décrivent, en détail, chaque
élément. Tout de l'endroit où ils peuvent aller à ce qui s'y passe est exposé de
manière facile à suivre."}]
```

6.4 Prediction for large dataset

[118]:
```
gen_kwargs = {"length_penalty": 0.8, "num_beams": 8, "max_length": 400}
custom_pipeline = pipeline(
    task="translation",
    model=checkpoint,
    tokenizer=checkpoint,
    device=0,
    framework = 'pt'
)
translations = custom_pipeline(test_data_flattened['en'], **gen_kwargs,␣
 ↪batch_size=16)
```

[119]:
```
translations
```

[119]:
```
[{'translation_text': 'Haut de la page & Y & #160;:'},
 {'translation_text': '-'},
 {'translation_text': 'Capitaliser le texte sélectionné ou le mot courant.'},
 {'translation_text': 'Ouvrir & konqueror;.'},
 {'translation_text': 'Chemins de recherche'},
 {'translation_text': 'Réduire à Grayscale'},
 {'translation_text': "L'utilisateur ne peut modifier aucun paramètre sur cette
page."},
```

```
 {'translation_text': 'Générer'},
 {'translation_text': 'Vietnamien (Vietnam)'},
 {'translation_text': 'Sélectionner le style à importer & #160;:'},
 {'translation_text': 'Cisco - Hub'},
 {'translation_text': "Impossible de lire l'en-tête de la base de données %1."},
 {'translation_text': 'Serveur trouvé, connexion…'},
 {'translation_text': 'Hôte'},
 {'translation_text': 'Envoyer un & fichier'},
 {'translation_text': 'toute la journée'},
 {'translation_text': 'exameneur; examenseur; Em'},
 {'translation_text': "L'alias est le nom que vous voulez donner à votre compte.
Il devrait être unique. Vous pouvez avoir plusieurs connexions au même service
de sorte que l'alias vous permet de leur donner des noms."},
 {'translation_text': 'Le & dataditor;'},
 {'translation_text': 'htdig'},
 {'translation_text': "Activer cette option si les messages importants ne
doivent jamais être supprimés pendant l'expiration du message, & cf; pendant la
suppression automatique des messages anciens."},
 {'translation_text': 'Verbeux'},
 {'translation_text': "Configurer une répétition d'alarme supplémentaire"},
 {'translation_text': "Le nombre d'octets lus. Voir ce que c'est pour plus
d'informations."},
 {'translation_text': 'Feuilles de tissu'},
 {'translation_text': "Sélectionnez la région dans laquelle vous voulez utiliser
les vacances ici. Les vacances définies sont affichées comme des jours non
ouvrables dans le navigateur de date, la vue de l'agenda, etc."},
 {'translation_text': 'Mason'},
 {'translation_text': 'Masque'},
 {'translation_text': 'Allemagne'},
 {'translation_text': 'Cliquez trois fois pour sélectionner la phrase
entière.'},
 {'translation_text': 'Chinois (PRC)'},
 {'translation_text': 'secondes'},
 {'translation_text': 'Cyan, Noir'},
 {'translation_text': 'Ne modifiez pas les données réelles du moteur.'},
 {'translation_text': "Impossible d'insérer la ligne parce qu'elle ne contient
pas la clé principale de la table maître entière."},
 {'translation_text': '& Marque de la branche & #160;:'},
 {'translation_text': "Système d'impression LPD & UNIX; générique (par
défaut)"},
 {'translation_text': 'BESSELK(3; 9) renvoie 397.95880'},
 {'translation_text': 'Supprimer'},
 {'translation_text': 'Modifier'},
 {'translation_text': "Insérez une image raster. Ceux-ci ne sont pas aussi
facilement graduables que les images vectorielles ou clipart. & kpresenter;
comprend actuellement de nombreux formats, y compris. tiff,. jpg,. png et bien
d'autres."},
```

```
 {'translation_text': 'Les mots de passe correspondent.'},
 {'translation_text': 'Fichiers de données espagnols & #160;: eXParTaKus
expartakus@expartakus. com'},
 {'translation_text': 'Taille du rect pour les mesures'},
 {'translation_text': "Vérifiez si le nom du fichier qui se trouve dans le texte
des éléments doit être analysé. Ceci n'a de sens que si l'élément peut contenir
un nom de fichier et si le fichierNameRx est spécifié."},
 {'translation_text': 'Ctrl; N Fichier'},
 {'translation_text': 'GUI XML'},
 {'translation_text': 'Détails'},
 {'translation_text': 'Cochez cette case si vous voulez rechercher des articles
Journal qui correspondent aux critères de recherche.'},
 {'translation_text': 'Validation de & XML;'},
 {'translation_text': '& Fichier'},
 {'translation_text': 'Création de widgets & kommander;'},
 {'translation_text': "Si cette case est cochée, Konqueror n'affichera que les
signets dans la barre d'outils du signet que vous avez marqué pour le faire dans
l'éditeur de signet."},
 {'translation_text': '& kaddressbook;'},
 {'translation_text': 'Mot de passe incorrect, veuillez réessayer.'},
 {'translation_text': 'Nom du champ'},
 {'translation_text': 'Configuration du système'},
 {'translation_text': 'Nicolas Ternisien'},
 {'translation_text': 'Internet Movie Database'},
 {'translation_text': 'Impossible de déterminer le système de partition de
fichiers du secteur %1 sur le périphérique %2.'},
 {'translation_text': "Sélectionner la vue de l'outil à ajouter"},
 {'translation_text': 'Définir si les fractions doivent être réduites.'},
 {'translation_text': "Réaction à l' invitation"},
 {'translation_text': 'Allemand (mise en page du clavier Neo 2)'},
 {'translation_text': 'Effacer la liste des signets.'},
 {'translation_text': 'Album de mon ami'},
 {'translation_text': 'Configuration du cadre pour %1'},
 {'translation_text': 'Containment'},
 {'translation_text': "Aucun indice de n'importe quelle couleur dans les tons
gris"},
 {'translation_text': "compare et fusionne deux ou trois fichiers ou répertoires
d'entrée de texte,"},
 {'translation_text': "Redimensionner l'image du titre %1 à %2x%3"},
 {'translation_text': 'Quitter facilement une application D-Bus activée'},
 {'translation_text': 'Max Blazejak'},
 {'translation_text': 'Nom de la session manquante'},
 {'translation_text': "Kate utilise deux formes différentes de modules, à savoir
les modules pour l'application & kate; et les modules pour le composant de
l'éditeur & kate;. Ce dernier est disponible pour toute application utilisant le
composant de l'éditeur, comme KDevelop, Quanta, Kile, Kwrite et bien d'autres,
tandis que les modules d'application sont spécifiques à l'application &
```

```
kate;."},
 {'translation_text': 'Faire défiler le message vers le bas'},
 {'translation_text': 'Enregistrer la liste des utilisateurs?'},
 {'translation_text': '%1 contacté. En attente de réponse…'},
 {'translation_text': 'Utilisez le fichier de configuration spécifié.'},
 {'translation_text': 'Longueur & #160;:'},
 {'translation_text': 'La boîte de dialogue & kpresenter; Create Memory Stick
Diaporama (extended).'},
 {'translation_text': 'Divers'},
 {'translation_text': 'Service de recherche'},
 {'translation_text': 'Réglage manuel de la vitesse du ventilateur. Ceci n\'est
efficace que lorsque la vitesse est réglée sur "Manuel".'},
 {'translation_text': "La connaissance des mathématiques ou de la langue n'est
pas nécessaire pour résoudre les énigmes de & kappname;."},
 {'translation_text': 'Stephan Binner'},
 {'translation_text': "Fournit une boîte de texte pour faciliter l'accès aux
moteurs de recherche comme Google."},
 {'translation_text': 'Nouveau point de sauvegarde'},
 {'translation_text': 'Canal vers le bas'},
 {'translation_text': 'Utilisez le chiffrement si le serveur le demande.'},
 {'translation_text': "C'est la liste des images à télécharger vers la cible
spécifiée."},
 {'translation_text': 'Documentation copyright 2002, 2003, 2005 & Anders. Lund;
& Anders. Lund. mail;'},
 {'translation_text': 'Si cela ne commence pas avec un /, alors il est supposé
être relatif à la racine du serveur. Le paramètre par défaut est / var/ log/
cups/ error_log.'},
 {'translation_text': 'Masse molaire du solvant'},
 {'translation_text': 'Aléatoire'},
 {'translation_text': 'min'},
 {'translation_text': 'Contrôle de la lecture (PBC)'},
 {'translation_text': 'Début tardif calculé & #160;: %1'},
 {'translation_text': "Le calcul de l'histogramme a échoué."},
 {'translation_text': 'Spraycan'},
 {'translation_text': 'Curseur Big LB'},
 {'translation_text': 'Basculer & Augmenter'},
 {'translation_text': 'Équipes de traduction KDE'},
 {'translation_text': 'bradh@frogmouth. net'},
 {'translation_text': 'H'},
 {'translation_text': '& Ctrl; Maj; B'},
 {'translation_text': '& Chercher'},
 {'translation_text': 'kdeaddons'},
 {'translation_text': 'RIDGE VERT'},
 {'translation_text': 'Choisir les contacts à imprimer'},
 {'translation_text': 'Marque & kruler; une longueur moyenne - environ 640
pixels de long.'},
 {'translation_text': "Si vous préférez vérifier l'orthographe explicitement,
```

```
plutôt que automatiquement, vous pouvez aussi forcer l'orthographe… vérifier
ici."},
 {'translation_text': "Utiliser le texte pour étiquetter les objets de la liste
d'observation"},
 {'translation_text': 'Réduire toutes les catégories'},
 {'translation_text': "Ce menu contient une option pour obtenir un indice (s'il
y en a un). Les jeux de championnat et de tutorial ont des indices à chaque
niveau. D'autres jeux ont rarement des indices. Il contient également des
options pour rejouer des enregistrements de jeu, comme lors de démonstrations,
montrant des exemples de solutions et rejouant vos propres efforts."},
 {'translation_text': 'Substrat'},
 {'translation_text': 'Bulqizë'},
 {'translation_text': "Erreur lors de l'ouverture du fichier %1. %2."},
 {'translation_text': "& Ouvrir l' emplacement"},
 {'translation_text': 'Ajouter une classe pour la génération de code'},
 {'translation_text': "Les options personnelles se rapportent à votre identité
personnelle, et d'autres choses diverses qui ne relèvent d'aucune des autres
catégories générales."},
 {'translation_text': 'Applique les paramètres actuels dans le module
actuellement ouvert.'},
 {'translation_text': 'Disque'},
 {'translation_text': 'c) 2001-2002 Alexander Neundorf'},
 {'translation_text': 'Sprite de Cavalry point supérieur gauche x-coordinate.
Ceci est donné relativement à la partie supérieure gauche de la carte,
dimensionné comme défini dans le widget de définition de la peau principale.
Cette valeur sera automatiquement mise à jour si vous déplacez la cavalerie.'},
 {'translation_text': 'Dvorak - ABCD'},
 {'translation_text': 'Avancé'},
 {'translation_text': "Calculer l'heure standard"},
 {'translation_text': 'Renommer le fichier, le dossier, etc.'},
 {'translation_text': 'Rendu indirect'},
 {'translation_text': 'Emplacement'},
 {'translation_text': "Exécution de l'état sur un élément externe à %1."},
 {'translation_text': '6 caractères'},
 {'translation_text': 'Hobart'},
 {'translation_text': 'Par défaut du projet'},
 {'translation_text': 'La boîte de dialogue de configuration de & bombardier;
vous permet de sélectionner des thèmes visuels pour & bombardier;.'},
 {'translation_text': '& Orientation'},
 {'translation_text': 'Coordonnée Y'},
 {'translation_text': 'Un gestionnaire de fenêtres keyboard-friendly avec des
fenêtres carrelées, basé sur PWM'},
 {'translation_text': 'Cisco - Interrupteur vocal'},
 {'translation_text': "Cette option spécifie le nombre d'espaces utilisés par
Tidy pour identifier le contenu, lorsque l'indentation est activée."},
 {'translation_text': 'Coordonnée & Grille'},
 {'translation_text': '& La première rangée contient des en-têtes'},
```

```
 {'translation_text': 'Bonne signature depuis & %1 ID de la clé & #160;: %2'},
 {'translation_text': 'Degrés Rankine'},
 {'translation_text': 'Cabano'},
 {'translation_text': 'Si cochée, les objets du catalogue Messier seront cachés
lorsque la carte sera en mouvement.'},
 {'translation_text': 'Tube RFB entrant'},
 {'translation_text': "Si vous utilisez un serveur de nouvelles local, les
articles n'apparaissent que si le serveur de nouvelles les a déjà obtenus depuis
Internet; si vous utilisez leafnode, cela se fait par le programme fetchnews."},
 {'translation_text': 'Lancement rapide'},
 {'translation_text': 'HP; JetDirect Protocol Direct TCP/ IP Printing'},
 {'translation_text': 'KDevelop GUI Builder basé sur Qt Designer'},
 {'translation_text': 'Nombre de cercles'},
 {'translation_text': 'Oklahoma'},
 {'translation_text': 'semaine'},
 {'translation_text': '& Charger la liste des chaînes depuis le fichier…'},
 {'translation_text': 'A partir du curseur'},
 {'translation_text': "Si les messages ne sont pas cochés, ils seront toujours
stockés non chiffrés dans votre dossier d'envoi, même s'ils sont cryptés."},
 {'translation_text': 'CRÉDIT_FOR_TRANSLATORS'},
 {'translation_text': "Sélectionne tout le texte dans la fenêtre de
l'éditeur."},
 {'translation_text': "La connexion peut ne pas contenir d'espaces ou de colons
!"},
 {'translation_text': 'Re_démarrer'},
 {'translation_text': 'Little Four Joys'},
 {'translation_text': "Choisissez le type de base de données que vous souhaitez
utiliser. La plupart des utilisateurs voudront choisir ici une simple base de
données locale. Cependant, vous pouvez également utiliser des serveurs distants
au moyen d'une base de données MySQL ou PostgreSQLTM."},
 {'translation_text': 'Moyenne'},
 {'translation_text': "La demande d'abonnement a échoué."},
 {'translation_text': 'Masse atomique'},
 {'translation_text': 'Exporte les certificats sélectionnés dans un fichier.'},
 {'translation_text': 'Vibrorg'},
 {'translation_text': 'Voile douches'},
 {'translation_text': 'Sortir du script'},
 {'translation_text': 'Écraser le fichier ?'},
 {'translation_text': 'Utiliser & stateful crypting'},
 {'translation_text': 'Périodes de neige humide'},
 {'translation_text': 'Activer le débogueur'},
 {'translation_text': 'La Massana'},
 {'translation_text': 'Accès aux applications cachées minimisé dans le plateau
du système'},
 {'translation_text': 'Auteur original'},
 {'translation_text': 'K'},
 {'translation_text': '& kappname;'},
```

```
 {'translation_text': "Utiliser l'affichage du serveur X 'affiche le nom '"},
 {'translation_text': "Environnement qui permet à LaTeX d'imprimer exactement ce
que vous saisissez. Dans cette variante, les espaces sont imprimés d'une manière
spéciale."},
 {'translation_text': 'Auteur'},
 {'translation_text': 'Compte préféré'},
 {'translation_text': "une entrée de marque d' arrêt"},
 {'translation_text': 'Ce filtre, maintenant, affiche tous les articles, pas
plus de 3 jours, contenant knode dans le sujet.'},
 {'translation_text': 'Changer le facteur de formulaire'},
 {'translation_text': 'Kevin Krammer'},
 {'translation_text': "Les appareils et programmes suivants n'ont pas pu être
mis à jour & #160;:"},
 {'translation_text': 'V'},
 {'translation_text': 'Heritage acls'},
 {'translation_text': 'Commentaire & #160;: %1'},
 {'translation_text': '*. mac_BAR_Maxima batch File'},
 {'translation_text': 'Impossible de télécharger le message daté %1 de %2 avec
le sujet %3 vers le serveur.'},
 {'translation_text': "Extension non valable, manquante'manifest. xml
'ou'source. tar. bz2'"},
 {'translation_text': 'ESSID & #160;:'},
 {'translation_text': 'Afficher à la taille réelle dans une vue différente.
Sinon, cliquez avec le bouton du milieu de la souris sur la bande dessinée.'},
 {'translation_text': 'Serveur MySQL trouvé & #160;: %1'},
 {'translation_text': '& Refaire'},
 {'translation_text': 'Solveur & #160;: ce jeu est perdu.'},
 {'translation_text': '& Démo principale'},
 {'translation_text': 'AppleSingle'},
 {'translation_text': 'Comment laisser des messages sur le serveur ?'},
 {'translation_text': 'SAO/ R96P'},
 {'translation_text': 'Si vous avez des plugins & konqueror; installés, il y
aura des entrées supplémentaires dans le menu Outils. Consultez le chapitre
Plugins de konqueror; pour plus de détails.'},
 {'translation_text': 'Vitesse de chargement assurée'},
 {'translation_text': 'Cacher les formulaires'},
 {'translation_text': 'Agrafer toutes les 12 feuilles'},
 {'translation_text': "KWord Filtre d'importation WordPerfect"},
 {'translation_text': 'Semences moyennes connectées par torrent de course'},
 {'translation_text': 'Créer un nouveau dossier de modèle'},
 {'translation_text': 'Quelques trucs et astuces'},
 {'translation_text': 'KDE'},
 {'translation_text': "Calculer l'expression"},
 {'translation_text': 'Ordre de tri par défaut'},
 {'translation_text': 'Insérer une boîte de texte'},
 {'translation_text': 'Affiche la présentation avec un visionneur & PostScript;,
exactement comme si elle était imprimée.'},
```

{'translation_text': 'Support pour de nombreuses villes en Slovaquie, mais avec des données statiques. Il y a aussi support pour bratislava avec des données dynamiques.'},
{'translation_text': 'Mirror Format est incorrect. Il doit être valide http, ftp, ou URL de fichier'},
{'translation_text': 'Propriétés du timbre'},
{'translation_text': 'Parallèlement'},
{'translation_text': "Cette dernière fonction mentionnée n'a pas été largement utilisée à l'intérieur de & kde; jusqu'à présent, car les développeurs ne sont pas encore pleinement conscients de la puissance de & kdeprint;. Attendez-vous à en avoir d'autres dans un avenir proche. Un exemple que j'ai découvert est l'application & kcron;. Il vous permet de modifier le crontab à travers un & GUI;. Les développeurs ont implémenté une fonction d'impression qui vous permet (ou la racine) de choisir si vous voulez imprimer l'ensemble du crontab (pour tous les utilisateurs) ou simplement la partie qui est marquée. Vous pouvez voir les effets sur & kdeprint; dans les captures d'écran suivantes."},
{'translation_text': 'Faire compter la note de marge sur le corps du texte'},
{'translation_text': 'Exporter vers le destinataire'},
{'translation_text': 'Rafraîchit les périphériques & #160;: cette commande force & partman; à analyser et à lire à nouveau les périphériques de votre ordinateur. Cela peut par exemple être utile si vous vous branchez sur un disque dur USB externe après avoir démarré & partman;.'},
{'translation_text': 'Fog/ brume de pluie légère'},
{'translation_text': 'Tous les bureaux'},
{'translation_text': 'KMathTrigGraphPart'},
{'translation_text': '& Chiffrer le message'},
{'translation_text': 'Exercice de conversion'},
{'translation_text': "Une fois le modem dans l'état de la commande, vous pouvez l'envoyer. Pour le raccrocher, envoyez la commande ATH. Les codes d'échappement et la chaîne d'attente utilisée par & kppp; sont affichés dans la boîte de dialogue Modem Commands. Ceux-ci doivent correspondre à votre modem."},
{'translation_text': 'O'},
{'translation_text': 'Copier toutes les pages (non triées)'},
{'translation_text': 'Std dev. & #160;:'},
{'translation_text': 'Éditable & #160;:'},
{'translation_text': 'Dollar des Îles Salomon'},
{'translation_text': 'Commandes à éditer'},
{'translation_text': 'Réseau, DNS, itinéraires, interfaces'},
{'translation_text': 'Luminosité'},
{'translation_text': 'Nauru'},
{'translation_text': 'Cadran vocal'},
{'translation_text': 'Taille du bouton & #160;:'},
{'translation_text': 'Critique'},
{'translation_text': 'Félicitations! Vous avez résolu le niveau %1 avec %2 moves! Avancer vers le suivant?'},
{'translation_text': 'Boîte de dialogue'},
{'translation_text': '… et la FAQ K3b ne vous aide pas…'},

```
 {'translation_text': 'Gilles Calier'},
 {'translation_text': 'En sélectionnant K Bouton Applications Configuration du
système.'},
 {'translation_text': 'Qualité à laquelle une carte animée est peinte.'},
 {'translation_text': 'Nouveaux sous-clés & #160;:'},
 {'translation_text': 'Impossible de créer une mise en page sans nom.'},
 {'translation_text': 'En cours'},
 {'translation_text': "Artiste de l'icône Kompare"},
 {'translation_text': "Activer/ Désactiver l' œil de l' oiseau"},
 {'translation_text': 'Tableau'},
 {'translation_text': 'Exécuter néanmoins'},
 {'translation_text': 'Velocity'},
 {'translation_text': 'Remarques & #160;:'},
 {'translation_text': 'Fichier et paramètres du programme.'},
 {'translation_text': "Pas de nom d' option donné"},
 {'translation_text': "Sélectionner l' année & Images"},
 {'translation_text': 'Dollar guyanais'},
 {'translation_text': '& knetwalk;, un jeu pour les administrateurs système.'},
 {'translation_text': "Menus définis par l' utilisateur"},
 {'translation_text': 'Définit un jour de semaine et une semaine spécifique dans
le mois sur lequel cet événement ou à faire doit se reproduire'},
 {'translation_text': 'Passwd discussion timeout & #160;:'},
 {'translation_text': '& Déplacement/ redimensionnement'},
 {'translation_text': '& Augmenter le volume'},
 {'translation_text': '0123456789'},
 {'translation_text': 'Pager'},
 {'translation_text': 'KImageMapEditor'},
 {'translation_text': 'Interruption'},
 {'translation_text': 'Folio américain'},
 {'translation_text': 'Le mot de passe que vous avez saisi a une faible force.
Pour améliorer la force du mot de passe, essayez & #160;: - en utilisant un mot
de passe plus long; - en utilisant un mélange de lettres majuscules et
minuscules; - en utilisant des nombres ou des symboles ainsi que des lettres.
Voulez-vous utiliser ce mot de passe de toute façon & #160;?'},
 {'translation_text': 'Ajouter un suivi'},
 {'translation_text': 'ou'},
 {'translation_text': "Si cochée, le clic droit sur un élément le sélectionnera,
mais ne désélectionnera pas d'autres éléments déjà sélectionnés."},
 {'translation_text': "Trouver la prochaine occurrence d'un morceau de texte
dans une présentation."},
 {'translation_text': 'Verbe'},
 {'translation_text': 'Configurer le partage de fichiers…'},
 {'translation_text': 'Impossible de créer des menus (peut-être %1 manquant ou
corrompu)'},
 {'translation_text': "Avertissement d'encodage de fichier"},
 {'translation_text': 'Aller à & #160;:'},
 {'translation_text': "Page d'exportation"},
```

```
 {'translation_text': 'Aliments importés'},
 {'translation_text': 'Afrique/ Conakry'},
 {'translation_text': 'Protocole de messagerie instantanée de KAddressbook'},
 {'translation_text': 'Tom Albers'},
 {'translation_text': 'Glossaire'},
 {'translation_text': "Charge la liste des bases de données depuis le serveur,
de sorte que vous pouvez en sélectionner une à l'aide de la boîte combo « Nom
»."},
 {'translation_text': 'Kugar'},
 {'translation_text': 'localErrorRatio'},
 {'translation_text': 'Chemin vers les sous-titres externes'},
 {'translation_text': 'Document à ouvrir'},
 {'translation_text': 'Ngultrum'},
 {'translation_text': 'Temps depuis la première séance de pratique en jours'},
 {'translation_text': '& Annuler'},
 {'translation_text': 'Description'},
 {'translation_text': 'Algorithme slider'},
 {'translation_text': "Manuel de l'utilisateur de & kde;"},
 {'translation_text': "Mise à jour de l'étiquette, de la branche ou de la
date"},
 {'translation_text': 'Stdin est vide, pas de travail envoyé.'},
 {'translation_text': "Re & #160;: c'est un test"},
 {'translation_text': 'Objets dans le ciel'},
 {'translation_text': "Basculer l'affichage des étiquettes de noms de
planètes"},
 {'translation_text': '& Signaler un bogue…'},
 {'translation_text': 'La bibliographie est une section en dehors de votre
document principal, et un exemple de code pour la bibliographie ressemblerait à
ce qui suit:'},
 {'translation_text': '160 kbs'},
 {'translation_text': 'Modifier'},
 {'translation_text': 'Secret & #160;: %1'},
 {'translation_text': 'Long Middle Short Cone Space (gonfle 32 bits/ canal)'},
 {'translation_text': "Information sur le droit d'auteur sur le mot croisé"},
 {'translation_text': 'Manuel & klettres;'},
 {'translation_text': "J'ai fait une bonne image et je veux la partager avec mes
amis. Puis-je l'enregistrer en tant qu'image ?"},
 {'translation_text': 'Tamil'},
 {'translation_text': 'AutoLoginDelay'},
 {'translation_text': 'CHAR(65) renvoie « A »'},
 {'translation_text': 'Vérifiez les jours de la semaine qui sont des jours de
travail'},
 {'translation_text': 'ISO A9'},
 {'translation_text': "Réglez la longueur de l'arc de votre tarte."},
 {'translation_text': 'Notification de nouveaux messages'},
 {'translation_text': "Le logiciel n'est pas en cours d'exécution"},
 {'translation_text': 'Choisissez simplement les conditions que vous souhaitez
```

```
inclure dans votre recherche, puis cliquez sur OK pour créer votre liste de
lecture de recherche!'},
 {'translation_text': 'Vue DUChain'},
 {'translation_text': 'Basculer vers le mode Affichage par défaut.'},
 {'translation_text': 'Cliquez ensuite sur le bouton flèche droite pour le
placer dans le panneau droit.'},
 {'translation_text': 'Apparence'},
 {'translation_text': 'Ce manuel décrit & kteatime;, le théière & kde;.'},
 {'translation_text': 'Adresse'},
 {'translation_text': 'retardAnnuler'},
 {'translation_text': "Avec ce champ d'entrée, vous spécifiez le dossier que
vous voulez charger pour créer le nouveau dictionnaire."},
 {'translation_text': 'Inverser'},
 {'translation_text': 'Crée une image RVB blanche de 1600 x 1200 pixels.'},
 {'translation_text': 'ROLES_OF_TRANSLATORS'},
 {'translation_text': 'RLE'},
 {'translation_text': 'Configuration des modèles'},
 {'translation_text': "Spécifier le comportement d'une session à choix
multiples"},
 {'translation_text': 'Définir le nom…'},
 {'translation_text': 'En décochant la case Sons, les effets sonores de &
clignote; seront supprimés.'},
 {'translation_text': 'Seuil & #160;:'},
 {'translation_text': 'Circuit - Capacitateur vertical'},
 {'translation_text': 'Aller'},
 {'translation_text': '& Fichier'},
 {'translation_text': 'Quitter & kguitar;'},
 {'translation_text': 'Propriété'},
 {'translation_text': 'Golf'},
 {'translation_text': "Abréviation pour S synchrone P eripheral O perations O n
L ine; SPOOL ing permet aux applications d'impression (et aux utilisateurs) de
poursuivre leur travail pendant que le travail est pris en charge par un démon
système, qui stocke le fichier à un emplacement temporaire jusqu'à ce que
l'imprimante soit prête à imprimer."},
 {'translation_text': 'Peter Hedlund'},
 {'translation_text': '… vous pouvez appuyer sur +, -, CTRL + ou CTRL - pour
changer rapidement les dates.'},
 {'translation_text': "Basculer l' orientation de fractionnement"},
 {'translation_text': '& kolourpaint; est un programme de peinture libre et
facile à utiliser pour & kde;.'},
 {'translation_text': 'Charger'},
 {'translation_text': 'Veuillez insérer un support DVD vide ou appendable dans
le lecteur%1'},
 {'translation_text': 'Menkent'},
 {'translation_text': 'Tous & #160;:'},
 {'translation_text': 'vide'},
 {'translation_text': 'Lorsque & kappname; démarre, vous êtes présenté avec un
```

```
menu principal.'},
 {'translation_text': 'Planifier la mise à jour de la demande'},
 {'translation_text': 'Construisez une courbe de Bézier quadratique avec ce
point de contrôle'},
 {'translation_text': '%1 rupiahs'},
 {'translation_text': 'Afrique/ Dar_es_Salaam'},
 {'translation_text': 'Préfixes numériques et chiffres romains'},
 {'translation_text': 'Écrire le lead-out peut prendre un certain temps.'},
 {'translation_text': 'Slider %1'},
 {'translation_text': 'Aperçu…'},
 {'translation_text': "Importation de certificat & #160;: le certificat CA n'a
pas pu être extrait."},
 {'translation_text': 'kuser'},
 {'translation_text': '192 kbs'},
 {'translation_text': "Si cette option est cochée, K3b vérifiera la
configuration du système pour tout problème au démarrage et lorsque
l'utilisateur change les paramètres."},
 {'translation_text': "L'ASAR BATTLEFIELD"},
 {'translation_text': 'Est native se déplaçant (possiblement wobbling windows)
active'},
 {'translation_text': 'Marge supérieure'},
 {'translation_text': 'Kile'},
 {'translation_text': 'Le menu Fichier'},
 {'translation_text': "& Solveur d'équation…"},
 {'translation_text': 'Développeur de base et auteur original'},
 {'translation_text': 'M3 est également courant, mais différent sur de
nombreuses marques'},
 {'translation_text': 'Nouvelle contrainte…'},
 {'translation_text': '& Pamela. Roberts;'},
 {'translation_text': 'Jujuy (JY)'},
 {'translation_text': "suffixe d'achèvement de Nick & #160;:"},
 {'translation_text': 'Ctrl; - Affichage'},
 {'translation_text': "& Paragraphe tiret par touches de barre d' outils &
#160;:"},
 {'translation_text': 'Jeu de minigolf KDE'},
 {'translation_text': 'Guadalajara'},
 {'translation_text': '& Album'},
 {'translation_text': 'et modifiez-le & #160;:'},
 {'translation_text': 'Flèche circulaire'},
 {'translation_text': '& Exporter le pilote…'},
 {'translation_text': '& Lauri. Watts; & Lauri. Watts. mail;'},
 {'translation_text': '4.400'},
 {'translation_text': 'MPEG Layer 2 (mp2)'},
 {'translation_text': 'Préserver la luminosité'},
 {'translation_text': '1.90'},
 {'translation_text': '(Défaut) Désactiver la prévention du vol pour XV'},
 {'translation_text': 'Choisissez comment combiner les options que vous
```

spécifiez ici avec les options par défaut définies dans les paramètres de KPlayer. Par défaut utilise les options des paramètres de KPlayer, définissez Pour les remplacer par les options définies ici, et ajoutez ces options aux options par défaut.'},
 {'translation_text': 'K3bSetup - modifier la permission de gravure sur CD/ DVD avec K3b'},
 {'translation_text': "Eva Brucherseifer pour l'écriture de KMath (kbruch; 's & GUI; est basé sur ce travail)"},
 {'translation_text': "Temps %1 & #160;: %2 & #160;: %3. %4 n'est pas valable."},
 {'translation_text': "Affiche la correspondance actuelle des lettres de réponse aux nombres d'indices."},
 {'translation_text': "Indique si les fichiers organisés renommeront des fichiers de sorte qu'ils ne contiennent que des caractères ASCII 7 bits."},
 {'translation_text': 'Clés de confiance disponibles'},
 {'translation_text': 'Dans les deux cas, & kspread; supposera que les données sont sous forme CSV et ouvrira une boîte de dialogue vous permettant de contrôler la manière dont les données sont extraites du fichier ou du presse-papiers et placées dans les cellules de la feuille de travail.'},
 {'translation_text': 'graphiques. xml'},
 {'translation_text': 'Impossible de supprimer le webseed %1, il fait partie du torrent.'},
 {'translation_text': 'Pour importer un fichier de données existant, saisissez son nom de fichier ici. Vous décrirez ensuite le contenu du fichier ci-dessous. Vous pouvez laisser le nom de fichier vide pour construire un fichier de catalogue avec un en-tête valide, mais pas de données.'},
 {'translation_text': "Barre d' outils principale"},
 {'translation_text': "Enregistrer les modifications d' objets"},
 {'translation_text': 'Jens Herden'},
 {'translation_text': 'Importation de certificats…'},
 {'translation_text': 'Renommer'},
 {'translation_text': 'chartreuse1'},
 {'translation_text': 'Dernière déclaration réconciliée & #160;: %1'},
 {'translation_text': 'Il semble que nous ayons été redirigés vers un autre serveur; je ne sais pas comment gérer cela.'},
 {'translation_text': 'Connexion'},
 {'translation_text': "Les options suivantes déterminent les aspects de base de l'opération:"},
 {'translation_text': 'Adresse Hw'},
 {'translation_text': 'Résistance du signal & #160;:'},
 {'translation_text': "La suppression des objets se fait en les sélectionnant d'abord, puis en faisant l'un ou l'autre de ces choix & #160;:"},
 {'translation_text': 'Chemin'},
 {'translation_text': 'Trouver des recettes correspondantes'},
 {'translation_text': '%1x%2x%3 brique, %4 mouvements de shuffling'},
 {'translation_text': "Les sections suivantes décrivent, en détail, chaque élément. Tout de l'endroit où ils peuvent aller à ce qui s'y passe est exposé de

manière facile à suivre."},
 {'translation_text': 'c), 2003 Cornelius Schumacher'},
 {'translation_text': 'Les actions montées seront affichées dans une vue de liste.'},
 {'translation_text': 'Comparaison de 2 fichiers & #160;:'},
 {'translation_text': 'Kjumpingcube'},
 {'translation_text': 'Message'},
 {'translation_text': "Pour déterminer si un fichier de signets donné a été fusionné dans les signets de l'utilisateur déjà ou non, & kbookmarkmerger; regarde simplement le nom du fichier de signets - le contenu n'est pas coché du tout. Cela signifie que changer un fichier de signets qui a déjà été fusionné dans les signets d'un utilisateur ne déclenchera pas la fusion une fois de plus."},
 {'translation_text': 'OpenOffice Polska LLC'},
 {'translation_text': 'République du Congo'},
 {'translation_text': "Permet aux entrées de la session actuelle d'être envoyées simultanément aux sessions sélectionnées par l'utilisateur (plus d'informations)"},
 {'translation_text': 'lignes'},
 {'translation_text': 'Pourrait déterminer le projet actif'},
 {'translation_text': 'Coller le contenu du presse-papiers…'},
 {'translation_text': 'Si cela est désactivé et que vous appuyez sur la touche tabulateur, un tabulaire est inséré, sinon le nombre approprié de caractères est inséré.'},
 {'translation_text': 'Active/ Désactive la construction de document ANT (Java)'},
 {'translation_text': 'Jaune'},
 {'translation_text': 'Créer une nouvelle collection'},
 {'translation_text': 'Le groupe auquel doit appartenir la socket de commande globale; peut être soit un nom, soit un identifiant numérique.'},
 {'translation_text': "Utilisez les paramètres de cache HTML à l'échelle de KDE lors du téléchargement des flux, afin d'éviter le trafic inutile. Désactivez seulement lorsque cela est nécessaire."},
 {'translation_text': 'KDE'},
 {'translation_text': 'Pluton'},
 {'translation_text': 'Inconnu'},
 {'translation_text': 'Configurer'},
 {'translation_text': 'Source'},
 {'translation_text': 'Envelopper les longues lignes'},
 {'translation_text': 'Le menu Fichier'},
 {'translation_text': "Si activé, le facteur de zoom sur un cadre n'est appliqué qu'au texte."},
 {'translation_text': 'Utilisateur & #160;:'},
 {'translation_text': "Le mimetype spécifié n'a pas d'usine de composants installée"},
 {'translation_text': "La séquence clé « & #160; %1 & #160; » est ambiguë. Utiliser & #160; « & #160; Configurer les raccourcis & #160; » depuis le menu «

& #160; Réglages & #160; » pour résoudre l'ambiguïté. Aucune action ne sera déclenchée."},
 {'translation_text': 'Il y a beaucoup de fonctionnalités que vous pouvez ajouter à votre document qui le rendront plus lisible et convivial. Vous pouvez ajouter des fonctionnalités telles que des polices spécifiques, des caractères gras, des italiques, souligner & etc; à votre document, et ces commandes se terminent par une commande \\\\end, ou à la fin de votre environnement.'},
 {'translation_text': '10.345 Lsh 3 = donne aussi 80.'},
 {'translation_text': 'Antonio Aloisio'},
 {'translation_text': "Ouvrir un fichier de base, puis déposer sa version modifiée dans la vue Sync primaire, puis Alt; Down ou Alt; Up (souvenez-vous que les raccourcis peuvent être modifiés d'une manière habituelle pour toutes les applications de & kde;) pour naviguer à travers les entrées qui sont différentes."},
 {'translation_text': 'Paraguay Guarani'},
 {'translation_text': 'Inverser/écraser le mode.'},
 {'translation_text': 'Argon'},
 {'translation_text': 'Limite de chargement'},
 {'translation_text': "Barre d' outils Stick"},
 {'translation_text': '*.svg_BAR_Scalable Vec tor Graphics'},
 {'translation_text': 'Janet'},
 {'translation_text': "Vos choix ici affectent chaque application & kde; qui rend HTML avec le propre rendu de & kde;, qui s'appelle khtml. Ceux-ci comprennent & kmail;, & khelpcenter; et bien sûr & konqueror;. Les choix ici n'affectent pas d'autres navigateurs comme & Netscape;."},
 {'translation_text': "Vous êtes sur le point de recréer l'index du dossier %1. Ceci détruira toutes les informations sur l'état du message. Voulez-vous vraiment continuer & #160;?"},
 {'translation_text': "Les raccourcis standard et les raccourcis globaux fonctionnent exactement de la même manière. En fait, d'une certaine manière, les raccourcis standard ou d'application sont aussi globaux. La seule différence est & #160;:"},
 {'translation_text': '& Largeur'},
 {'translation_text': "HotSync est désactivé parce que KPilot n'a pas pu déterminer l'état de l'économiseur d'écran. Vous pouvez désactiver cette fonction de sécurité en décochant la 'ne pas synchroniser lorsque l'économiseur d'écran est actif' dans la page HotSync de la boîte de dialogue de configuration."},
 {'translation_text': 'Un widget de boîte combo'},
 {'translation_text': "Mes articles n'apparaissent pas dans le groupe de discussion."},
 {'translation_text': '& Pipe vers le terminal'},
 {'translation_text': 'Tailles spécifiées sont incohérentes!'},
 {'translation_text': 'IP & #160;:'},
 {'translation_text': 'Afficher le nombre de & lignes'},
 {'translation_text': "Utilisation de la barre d' outils & #160;: cliquez sur l'icône"},

```
 {'translation_text': '(C) 2005 - 2009, Les développeurs KGet (C) 2001 - 2002,
Patrick Charbonnier (C) 2002, Carsten Pfeiffer (C) 1998 - 2000, Matej Koss'},
 {'translation_text': 'sous-sous-section: une section secondaire entre le sous-
paragraphe et le paragraphe.'},
 {'translation_text': "Dans vous pouvez voir la boîte de dialogue qui vous aide
à importer un projet en tant que module. Pour accéder à la boîte de dialogue
d'importation de & cervisia;, choisissez l'élément du menu Importer… du
dépôt."},
 {'translation_text': 'Bevel & #160;:'},
 {'translation_text': "Albums sans couverture qui n'affiche que les albums qui
n'ont pas d'images de couverture."},
 {'translation_text': 'pas de configuration de version (erreur de
programmeur)'},
 {'translation_text': 'Oui, yeux rouges, lumière de retour'},
 {'translation_text': "D'autres vies peuvent être gagnées en attrapant le don de
vie supplémentaire."},
 {'translation_text': "Impossible d'utiliser le modificateur final sur un membre
de classe abstrait."},
 {'translation_text': 'Ajouter aux listes de lecture'},
 {'translation_text': 'Utiliser une autre couleur de fond'},
 {'translation_text': 'Crédits'},
 {'translation_text': "Vous avez besoin d'une variable et d'une valeur ou d'une
variable pour faire un '='"},
 {'translation_text': 'Adresse commerciale Rue'},
 {'translation_text': 'Retyper le mot de passe & #160;:'},
 {'translation_text': 'Option de menu de fichiers pour créer un diaporama
HTML…'},
 {'translation_text': 'Afficher le texte défilant'},
 {'translation_text': 'Porto Rico (USA)'},
 {'translation_text': 'Hauteur & #160;:'},
 {'translation_text': 'Pause le jeu'},
 {'translation_text': 'État & #160;: %1'},
 {'translation_text': "Si la case Activer l'anti-aliasing est cochée, la méthode
anti-aliasing non récursive sera utilisée pour rendre la scène. Vous pouvez
configurer les valeurs de profondeur et de seuil. Voir la section Modes de rendu
pour une description détaillée des paramètres."},
 {'translation_text': 'Erreur'},
 {'translation_text': "Je ne peux pas faire un '< =' sans 2 variables"},
 {'translation_text': 'Aller sur le site Web de KDE'}]
```

[ ]: