English-to-French Data Preprocessing (Tensor Preparation)

```
# Step 1: Download and Load Dataset
import os
import tensorflow as tf
import numpy as np
import re
import requests

url = "http://www.manythings.org/anki/fra-eng.zip"
zip_path = tf.keras.utils.get_file("fra-eng.zip", origin=url, extract=True)
file_path = os.path.dirname(zip_path) + "/fra.txt"

with open(file_path, "r", encoding="utf-8") as f:
    lines = f.read().split("\n")

n_samples = 10000
word_pairs = [line.split('\t') for line in lines[:n_samples]]

# Step 2: Preprocess Sentences
def preprocess_sentence(sentence):
    sentence = sentence.lower().strip()
    sentence = re.sub(r"([?.!,¿])", r" \1 ", sentence)
    sentence = re.sub(r'[" "]+', " ", sentence)
    sentence = re.sub(r"[^a-zA-Z?.!,¿]+", " ", sentence)
    sentence = sentence.strip()
    sentence = '<start> ' + sentence + ' <end>'
    return sentence

input_texts = [preprocess_sentence(en) for en, fr in word_pairs]
target_texts = [preprocess_sentence(fr) for en, fr in word_pairs]

# Step 3: Tokenize and Pad Sequences
inp_tokenizer = tf.keras.preprocessing.text.Tokenizer(filters='')
inp_tokenizer.fit_on_texts(input_texts)
input_tensor = inp_tokenizer.texts_to_sequences(input_texts)
input_tensor        =        tf.keras.preprocessing.sequence.pad_sequences(input_tensor,
padding='post')

targ_tokenizer = tf.keras.preprocessing.text.Tokenizer(filters='')
targ_tokenizer.fit_on_texts(target_texts)
target_tensor = targ_tokenizer.texts_to_sequences(target_texts)
target_tensor       =       tf.keras.preprocessing.sequence.pad_sequences(target_tensor,
padding='post')

# Step 4: Check Shapes
print(f"Input tensor shape: {input_tensor.shape}")
print(f"Target tensor shape: {target_tensor.shape}")
```