```
#4. POS-taggers (30% for BS students – 20% for MS students)

import stanza
from collections import Counter
import pickle

# For Norwegian
nlp = stanza.Pipeline(lang="no", processors="tokenize,pos")

# Load corpus
with open("elo_tesla.txt", "r", encoding="utf-8") as file:
    corpus = file.read()

# Process first 20thousand words
doc = nlp(corpus[:20000])

# Extract tokens and POS tags
tokens_with_pos = [(word.text, word.upos) for sentence in doc.sentences for word in
sentence.words]

# save tagged tokens to reuse
with open("tagged_text.pkl", "wb") as pkl_file:
    pickle.dump(tokens_with_pos, pkl_file)

# a) Frequency of all unique POS
def pos_frequency(tagged_tokens):
    pos_tags = [pos for _, pos in tagged_tokens]
    return Counter(pos_tags).most_common(10)

# b) Words with more than one tag
def words_with_multiple_tags(tagged_tokens):
    tag_dict = {}
    for word, tag in tagged_tokens:
        if word not in tag_dict:
            tag_dict[word] = set()
        tag_dict[word].add(tag)

    multi_tag_words = {word: tags for word, tags in tag_dict.items() if len(tags) >
1}
    top_multi_tag_words = Counter({word: len(tags) for word, tags in
multi_tag_words.items()}).most_common(10)
    return len(multi_tag_words), top_multi_tag_words

# c) Word with the most tags
def word_with_most_tags(tagged_tokens):
    tag_dict = {}
    for word, tag in tagged_tokens:
        if word not in tag_dict:
            tag_dict[word] = set()
        tag_dict[word].add(tag)

    max_tags = max(len(tags) for tags in tag_dict.values())
    most_tagged_words = [word for word, tags in tag_dict.items() if len(tags) ==
max_tags]
    return most_tagged_words, max_tags

# d) Frequency of top 10 most common word pairs
def top_word_tag_pairs(tagged_tokens):
    return Counter(tagged_tokens).most_common(10)
```

```
# Run functiong
pos_freq = pos_frequency(tokens_with_pos)
multi_tag_words_count, top_multi_tag_words =
words_with_multiple_tags(tokens_with_pos)
most_tagged_words, max_tags = word_with_most_tags(tokens_with_pos)
common_word_tag_pairs = top_word_tag_pairs(tokens_with_pos)

# Print
print("a) Top 10 POS-tag frequencies:", pos_freq)
print("\nb) Words with more than one tag:")
print(f"Count: {multi_tag_words_count}, Top 10: {top_multi_tag_words}")
print("\nc) Word(s) with the most tags:", most_tagged_words, f"({max_tags} tags)")
print("\nd) Top 10 most common word-tag pairs:", common_word_tag_pairs)
```