

Test de mi-semester (Semestre 1)
Linguistique de corpus en anglais

- 1) Give the definition of a corpus.
- 2) Why is it a good idea to use corpora to conduct analyses or create algorithms in NLP?
- 3) Give 3 examples of tasks in NLP.
- 4) Write the 3 commands you would use in your terminal to
 - a. create a directory called MyProject
 - b. change directories to this newly created one
 - c. create a python file called NLP_task.py
- 5) Try to explain why it is a good idea to create a virtual environment for each of your projects.
- 6) What is a token ? Give some examples.
- 7) Write a small function in python which allows you to tokenize a sentence such as “*You shall know a word by the company it keeps.*” The function should return a list of tokens, ideally without any ‘\n’ characters.
- 8) Write lines in python which allow you to
 - a. open a file called “MyCorpus.txt”.
 - b. dump its contents into a variable called my_corpus (the variable’s type can be list or string).
- 9) Given the event *Rolling 1 six-sided die (dé)*, what is the probability of
 - a. Rolling a 1
 - b. Rolling a 2 or a 6 ?
 - c. Rolling an even number ?
- 10) You pick a card at random from a regular deck of 52 cards with 4 suits (clubs, diamonds, hearts, spades). Each suit has 13 cards: 2-10, ace, jack, queen, king.
 - a. What is the probability of picking a jack ? P (Jack)
 - b. What is the probability of picking hearts ? P (Heart)
 - c. What is the probability of picking the jack of hearts ? P (Jack and Heart)
 - d. What is the probability of picking a jack or hearts ? P (Jack or Heart)