

```
In [1]: import nltk
```

```
In [2]: nltk.download() # 전체 다운로드 시, 3GB이상의 대용량이므로 시간이 소요됨.
```

```
showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml (https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml)
```

```
showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml (https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml)
```

```
Out[2]: True
```

- all 패키지를 선택 후, 다운로드 진행.

NLTK Downloader

File View Sort Help

Collections Corpora Models All Packages

Identifier	Name	Size	Status
all	All packages	n/a	partial
all-corpora	All the corpora	n/a	partial
all-nltk	All packages available on nltk_data gh-pages branch	n/a	partial
book	Everything used in the NLTK Book	n/a	partial
popular	Popular packages	n/a	partial
tests	Packages for running tests	n/a	partial
third-party	Third-party data packages	n/a	not installed

Download Refresh

Server Index: https://raw.githubusercontent.com/nltk/nltk_data/g

Download Directory: C:\Users\WITHJS\AppData\Roaming\nltk_data

```
In [5]: from nltk.corpus import brown  
brown.words()
```

```
Out[5]: ['The', 'Fulton', 'County', 'Grand', 'Jury', 'said', ...]
```

NLTK 데이터 다운로드 위치

- NLTK 다운로드 윈도우에서 다운로드 디렉터리를 표준 공용 디렉터리로 설정하기 위해 다음과 같이 설정
 - 윈도우 : C:\nltk_data
 - Linux : /usr/share/nltk_data
 - Mac : /usr/local/share/nltk_data
- 만약 표준 디렉터리를 설정하지 않았을 경우, 환경 변수를 이용하여 설정한다.

01. 토큰화

```
In [7]: import nltk  
sentence = """At eight o'clock on Thursday morning  
Arthur didn't feel very good."""  
  
tokens = nltk.word_tokenize(sentence)  
print(tokens)
```

```
['At', 'eight', 'o'clock', 'on', 'Thursday', 'morning', 'Arthur', 'did', 'n't', 'feel', 'very', 'good', '.']
```

02. 품사 태깅

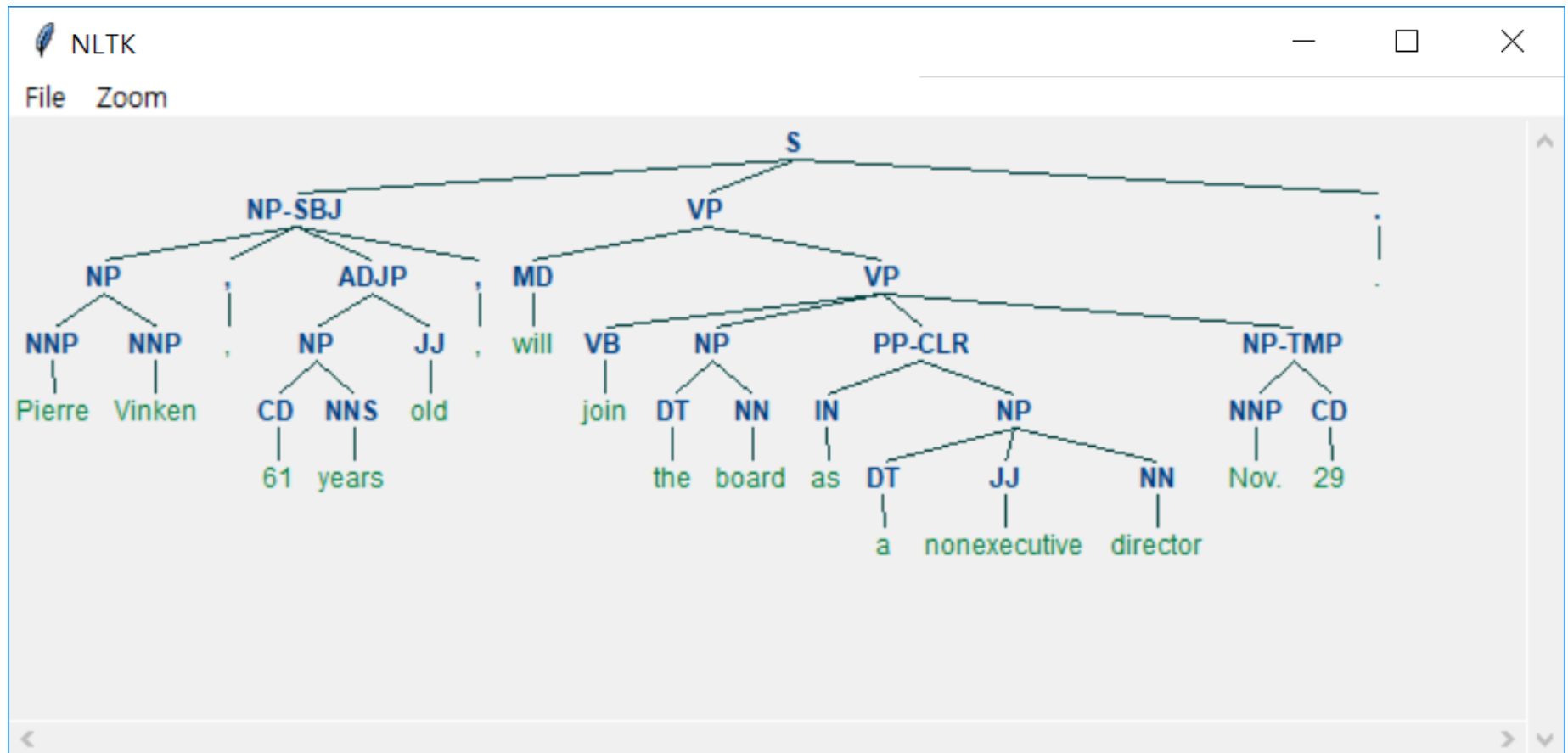
```
In [8]: tagged = nltk.pos_tag(tokens)
tagged[0:6]
```

```
Out[8]: [('At', 'IN'),
         ('eight', 'CD'),
         ("o'clock", 'NN'),
         ('on', 'IN'),
         ('Thursday', 'NNP'),
         ('morning', 'NN')]
```

- NN : Noun(명사)
- IN : preposition(전치사)
- CD : cardinal numbers (기수)
- NNP : proper noun, singular (고유명사)
- 기타 리스트 참조 : <https://pythonprogramming.net/part-of-speech-tagging-nltk-tutorial/> (<https://pythonprogramming.net/part-of-speech-tagging-nltk-tutorial/>)

03. 구문 분석 트리 출력

```
In [11]: from nltk.corpus import treebank
t = treebank.parsed_sents('wsj_0001.mrg')[0]
t.draw()
```



In []: