

NLP

Find your favorite news source and grab the article text.

1. Show the most common words in the article.
2. Show the most common words under a part of speech. (i.e. NOUN: {'Bob':12, 'Alice':4,})
3. Find a subject/object relationship through the dependency parser in any sentence.
4. Show the most common Entities and their types.
5. Find Entities and their dependency (hint: entity.root.head)
6. Find the most similar words in the article

Note: Yes, the notebook from the video is not provided, I leave it to you to make your own :) it's your final assignment for the semester.
Enjoy!

```
In [ ]: # Text taken from Nature article: https://www.nature.com/articles/d41586-024-01103-7
# doi: https://doi.org/10.1038/d41586-024-01103-7
# Title: Peter Higgs obituary: physicist who predicted boson that explains why particles have mass
# Author: Christine Sutton
# Date: 12 April 2024

text = """During a few weeks in the summer of 1964, Peter Higgs, a theoretical physicist at the University of Edinburgh The mechanism Higgs described became a key component of the standard model of particle physics during the 1970s, but the Higgs was born in Newcastle upon Tyne in 1929, but his father's work as a sound engineer for the BBC took the family to As a hitch-hiking student, Higgs had discovered a liking for Edinburgh, so in 1960 he was happy to be appointed as a lec This seemed a dead end, but in 1964, Higgs realized that it was possible to get round the difficulty using gauge theory The field had seemed something of a scientific backwater in the 1960s, but Higgs was not working alone. Two other papers Around this time, after his marriage broke down, Higgs found he was losing his way in theoretical particle physics. In a The award of the Nobel prize to Higgs and Belgian physicist François Englert – the surviving author of the paper publish A keen music lover, Higgs had little interest in the trappings of modern technology – he famously had no television and
```

Download the model from SpaCy

```
In [ ]: !python -m spacy download en_core_web_lg
```

```
Collecting en-core-web-lg==3.7.1
  Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web_lg-3.7.1/en_core_web_lg-3.7.1-py3
  -none-any.whl (587.7 MB)
----- 0.0/587.7 MB ? eta -:--:-
----- 0.0/587.7 MB ? eta -:--:-
----- 0.0/587.7 MB 262.6 kB/s eta 0:37:19
----- 0.1/587.7 MB 409.6 kB/s eta 0:23:55
----- 0.4/587.7 MB 1.9 MB/s eta 0:05:16
----- 1.5/587.7 MB 6.7 MB/s eta 0:01:28
----- 3.4/587.7 MB 12.7 MB/s eta 0:00:46
----- 5.8/587.7 MB 18.6 MB/s eta 0:00:32
----- 7.9/587.7 MB 21.1 MB/s eta 0:00:28
----- 10.2/587.7 MB 25.2 MB/s eta 0:00:23
----- 13.0/587.7 MB 46.9 MB/s eta 0:00:13
----- 16.7/587.7 MB 65.2 MB/s eta 0:00:09
----- 21.5/587.7 MB 93.0 MB/s eta 0:00:07
----- 27.1/587.7 MB 131.2 MB/s eta 0:00:05
----- 31.4/587.7 MB 110.0 MB/s eta 0:00:06
----- 35.7/587.7 MB 93.9 MB/s eta 0:00:06
----- 38.6/587.7 MB 81.8 MB/s eta 0:00:07
----- 43.1/587.7 MB 81.8 MB/s eta 0:00:07
----- 47.2/587.7 MB 93.9 MB/s eta 0:00:06
----- 50.9/587.7 MB 93.9 MB/s eta 0:00:06
----- 54.7/587.7 MB 93.9 MB/s eta 0:00:06
----- 57.8/587.7 MB 93.9 MB/s eta 0:00:06
----- 61.0/587.7 MB 81.8 MB/s eta 0:00:07
----- 64.3/587.7 MB 72.6 MB/s eta 0:00:08
----- 68.3/587.7 MB 81.8 MB/s eta 0:00:07
----- 73.1/587.7 MB 108.8 MB/s eta 0:00:05
----- 78.1/587.7 MB 108.8 MB/s eta 0:00:05
----- 82.3/587.7 MB 108.8 MB/s eta 0:00:05
----- 87.7/587.7 MB 108.8 MB/s eta 0:00:05
----- 92.4/587.7 MB 131.2 MB/s eta 0:00:04
----- 97.1/587.7 MB 108.8 MB/s eta 0:00:05
----- 101.6/587.7 MB 108.8 MB/s eta 0:00:05
----- 104.9/587.7 MB 108.8 MB/s eta 0:00:05
----- 110.2/587.7 MB 108.8 MB/s eta 0:00:05
----- 114.2/587.7 MB 93.0 MB/s eta 0:00:06
----- 118.9/587.7 MB 131.2 MB/s eta 0:00:04
----- 125.1/587.7 MB 131.2 MB/s eta 0:00:04
----- 130.1/587.7 MB 165.0 MB/s eta 0:00:03
----- 135.2/587.7 MB 131.2 MB/s eta 0:00:04
----- 141.2/587.7 MB 165.0 MB/s eta 0:00:03
----- 147.6/587.7 MB 165.0 MB/s eta 0:00:03
----- 153.0/587.7 MB 165.0 MB/s eta 0:00:03
```

```
----- 158.6/587.7 MB 165.0 MB/s eta 0:00:03
----- 163.5/587.7 MB 131.2 MB/s eta 0:00:04
----- 168.8/587.7 MB 131.2 MB/s eta 0:00:04
----- 173.6/587.7 MB 129.5 MB/s eta 0:00:04
----- 178.8/587.7 MB 131.2 MB/s eta 0:00:04
----- 183.6/587.7 MB 131.2 MB/s eta 0:00:04
----- 190.2/587.7 MB 162.4 MB/s eta 0:00:03
----- 194.4/587.7 MB 131.2 MB/s eta 0:00:03
----- 202.0/587.7 MB 162.4 MB/s eta 0:00:03
----- 208.4/587.7 MB 162.4 MB/s eta 0:00:03
----- 213.2/587.7 MB 131.2 MB/s eta 0:00:03
----- 219.7/587.7 MB 162.4 MB/s eta 0:00:03
----- 225.3/587.7 MB 162.4 MB/s eta 0:00:03
----- 231.5/587.7 MB 131.2 MB/s eta 0:00:03
----- 239.2/587.7 MB 162.4 MB/s eta 0:00:03
----- 246.4/587.7 MB 162.4 MB/s eta 0:00:03
----- 253.5/587.7 MB 165.0 MB/s eta 0:00:03
----- 259.4/587.7 MB 162.4 MB/s eta 0:00:03
----- 264.1/587.7 MB 162.4 MB/s eta 0:00:02
----- 271.9/587.7 MB 162.4 MB/s eta 0:00:02
----- 278.6/587.7 MB 162.4 MB/s eta 0:00:02
----- 286.0/587.7 MB 217.7 MB/s eta 0:00:02
----- 293.3/587.7 MB 162.4 MB/s eta 0:00:02
----- 301.1/587.7 MB 162.4 MB/s eta 0:00:02
----- 308.3/587.7 MB 162.4 MB/s eta 0:00:02
----- 317.0/587.7 MB 217.7 MB/s eta 0:00:02
----- 324.2/587.7 MB 131.2 MB/s eta 0:00:03
----- 328.4/587.7 MB 129.5 MB/s eta 0:00:03
----- 334.1/587.7 MB 110.0 MB/s eta 0:00:03
----- 341.1/587.7 MB 217.7 MB/s eta 0:00:02
----- 347.1/587.7 MB 162.4 MB/s eta 0:00:02
----- 353.6/587.7 MB 131.2 MB/s eta 0:00:02
----- 359.9/587.7 MB 162.4 MB/s eta 0:00:02
----- 364.9/587.7 MB 129.5 MB/s eta 0:00:02
----- 370.2/587.7 MB 110.0 MB/s eta 0:00:02
----- 375.4/587.7 MB 131.2 MB/s eta 0:00:02
----- 380.8/587.7 MB 129.5 MB/s eta 0:00:02
----- 385.3/587.7 MB 131.2 MB/s eta 0:00:02
----- 389.8/587.7 MB 108.8 MB/s eta 0:00:02
----- 394.5/587.7 MB 108.8 MB/s eta 0:00:02
----- 400.7/587.7 MB 110.0 MB/s eta 0:00:02
----- 407.0/587.7 MB 131.2 MB/s eta 0:00:02
----- 412.3/587.7 MB 129.5 MB/s eta 0:00:02
----- 417.2/587.7 MB 131.2 MB/s eta 0:00:02
----- 422.4/587.7 MB 131.2 MB/s eta 0:00:02
```



```
0,>=3.7.2->en-core-web-lg==3.7.1) (68.0.0)
Requirement already satisfied: packaging>=20.0 in c:\users\jomors\appdata\local\anaconda3\lib\site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-lg==3.7.1) (23.1)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in c:\users\jomors\appdata\local\anaconda3\lib\site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-lg==3.7.1) (3.3.0)
Requirement already satisfied: numpy>=1.19.0 in c:\users\jomors\appdata\local\anaconda3\lib\site-packages (from spacy<3.8.0,>=3.7.2->en-core-web-lg==3.7.1) (1.24.3)
Requirement already satisfied: typing-extensions>=4.2.0 in c:\users\jomors\appdata\local\anaconda3\lib\site-packages (from pydantic!=1.8,!>=1.8.1,<3.0.0,>=1.7.4->spacy<3.8.0,>=3.7.2->en-core-web-lg==3.7.1) (4.7.1)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\jomors\appdata\local\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy<3.8.0,>=3.7.2->en-core-web-lg==3.7.1) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\jomors\appdata\local\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy<3.8.0,>=3.7.2->en-core-web-lg==3.7.1) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\jomors\appdata\local\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy<3.8.0,>=3.7.2->en-core-web-lg==3.7.1) (1.26.16)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\jomors\appdata\local\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy<3.8.0,>=3.7.2->en-core-web-lg==3.7.1) (2024.2.2)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in c:\users\jomors\appdata\local\anaconda3\lib\site-packages (from thinc<8.3.0,>=8.1.8->spacy<3.8.0,>=3.7.2->en-core-web-lg==3.7.1) (0.7.9)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in c:\users\jomors\appdata\local\anaconda3\lib\site-packages (from thinc<8.3.0,>=8.1.8->spacy<3.8.0,>=3.7.2->en-core-web-lg==3.7.1) (0.1.4)
Requirement already satisfied: colorama in c:\users\jomors\appdata\local\anaconda3\lib\site-packages (from tqdm<5.0.0,>=4.38.0->spacy<3.8.0,>=3.7.2->en-core-web-lg==3.7.1) (0.4.6)
Requirement already satisfied: click<9.0.0,>=7.1.1 in c:\users\jomors\appdata\local\anaconda3\lib\site-packages (from typer<0.10.0,>=0.3.0->spacy<3.8.0,>=3.7.2->en-core-web-lg==3.7.1) (8.0.4)
Requirement already satisfied: cloudpathlib<0.17.0,>=0.7.0 in c:\users\jomors\appdata\local\anaconda3\lib\site-packages (from weasel<0.4.0,>=0.1.0->spacy<3.8.0,>=3.7.2->en-core-web-lg==3.7.1) (0.16.0)
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\jomors\appdata\local\anaconda3\lib\site-packages (from jinja2->spacy<3.8.0,>=3.7.2->en-core-web-lg==3.7.1) (2.1.1)
✓ Download and installation successful
You can now load the package via spacy.load('en_core_web_lg')
```

Set up a text processor and load the provided news article.

```
In [ ]: import spacy
```

```
In [ ]: nlp = spacy.load('en_core_web_lg')
doc = nlp(text)
doc
```

Out[]: During a few weeks in the summer of 1964, Peter Higgs, a theoretical physicist at the University of Edinburgh, UK, wrote two short papers outlining his ideas for a mechanism that could give mass to fundamental particles, the building blocks of the Universe. His aim was to rescue a theory that was mathematically appealing but ultimately unrealistic because the particles it described had no mass. The second paper drew attention to a measurable consequence of his proposal – it predicted the existence of a new massive particle. Nearly half a century later, the discovery of the predicted particle brought Higgs, who has died aged 94, a share of the 2013 Nobel Prize in Physics.

The mechanism Higgs described became a key component of the standard model of particle physics during the 1970s, but the associated particle remained stubbornly elusive. Then, in 2012, two giant experiments run by more than 6,000 physicists at CERN, Europe's high-energy physics laboratory near Geneva, Switzerland, discovered something with the appropriate properties. By then, the particle had achieved fame as the Higgs boson, although the self-effacing Higgs would usually refer to it as ‘the scalar boson’ in reference to its key characteristic of having no intrinsic spin.

Higgs was born in Newcastle upon Tyne in 1929, but his father’s work as a sound engineer for the BBC took the family to Bristol, where Higgs attended Cotham Grammar School. There, he spotted several mentions on the honours boards of a previous pupil, Paul Dirac, who had earned a share of the 1933 Nobel Prize in Physics for his work in quantum mechanics. Inspired, Higgs took up physics at King’s College London, where he obtained a PhD in 1954.

As a hitch-hiking student, Higgs had discovered a liking for Edinburgh, so in 1960 he was happy to be appointed as a lecturer there. He picked up on a long-standing interest in symmetry in subatomic particle physics, inspired in particular by the work of the future Nobel prizewinner Yoichiro Nambu, a Japanese American physicist then at the University of Chicago in Illinois. In physics, symmetry is linked to the conservation of quantities such as energy, momentum and electric charge. Working on a theory that had an underlying symmetry but in which particles had no mass, Nambu was attempting to generate mass through a mechanism known as spontaneous symmetry breaking. However, such symmetry breaking would also produce massless particles with zero spin, for which there was no evidence.

This seemed a dead end, but in 1964, Higgs realized that it was possible to get round the difficulty using gauge theory, which has the kind of symmetry found, for example, in the established theory of electromagnetism. Higgs showed that the massless particles associated with spontaneous symmetry breaking become ‘absorbed’ into massive particles. He published two short papers on this theme^{1,2}, the second of which explicitly predicted a massive spin-zero particle. It fell to other physicists to realize that this mechanism for spontaneous symmetry breaking was key to a mathematically coherent gauge theory of particle physics that unites the electromagnetic interactions between particles with the weak interactions involved in certain forms of radioactivity. This Nobel-prizewinning ‘electro-weak theory’ became one of the twin pillars of the standard model of particle physics, and was well established by experiments by the 1990s.

The field had seemed something of a scientific backwater in the 1960s, but Higgs was not working alone. Two other papers were published in 1964 on the mechanism^{3,4}, one appearing just before his own. But only Higgs drew attention to the associated massive spin-zero particle, which in the 1970s began to be called the Higgs boson. The catchy name stuck.

Around this time, after his marriage broke down, Higgs found he was losing his way in theoretical particle physics. In addition to teaching, he became involved in the union side of university life and wrote few physics papers. Nevertheless, as appreciation of his work on spontaneous symmetry breaking grew, he was increasingly asked to give talks, a popular title being ‘My life as a boson’. He retired in 1996.

The award of the Nobel prize to Higgs and Belgian physicist François Englert – the surviving author of the paper published just before Higgs’s in 1964 – came as little surprise to anyone, including Higgs, because it had been mooted since the 1980s. However, interest in the Higgs boson had increased over the years, not only among particle physicists but also among the general public and the media. It reached fever pitch after the construction of CERN’s Large Hadron Collider, which was billed by many as the machine that would discover the last missing piece of the standard model. This intense interest brought fame at a level no one could have imagined in the 1960s, and the quiet physicist became a media star, before retiring for a second time when he reached 85.

A keen music lover, Higgs had little interest in the trappings of modern technology – he famously had no television and

did not use the Internet. Yet he was far from being remote from the world. He had a strong social conscience and was a member of the Campaign for Nuclear Disarmament and Greenpeace at various times. Humble in many respects, with an infectious sense of humour, he was proud of the work that he had always known was important and which ultimately brought him fame.

Most common words

```
In [ ]: # Create an array of tokens from the text
words = [token.text                      # Get the text of the token, Load into an array
         for token in doc                 # Iterate over the tokens
         if not token.is_punct]          # Remove stop words and punctuation
print(words)
```

['During', 'a', 'few', 'weeks', 'in', 'the', 'summer', 'of', '1964', 'Peter', 'Higgs', 'a', 'theoretical', 'physicist', 'at', 'the', 'University', 'of', 'Edinburgh', 'UK', 'wrote', 'two', 'short', 'papers', 'outlining', 'his', 'ideas', 'for', 'a', 'mechanism', 'that', 'could', 'give', 'mass', 'to', 'fundamental', 'particles', 'the', 'building', 'blocks', 'of', 'the', 'Universe', 'His', 'aim', 'was', 'to', 'rescue', 'a', 'theory', 'that', 'was', 'mathematically', 'appealing', 'but', 'ultimately', 'unrealistic', 'because', 'the', 'particles', 'it', 'described', 'had', 'no', 'mass', 'The', 'second', 'paper', 'drew', 'attention', 'to', 'a', 'measurable', 'consequence', 'of', 'his', 'proposal', 'it', 'predicted', 'the', 'existence', 'of', 'a', 'new', 'massive', 'particle', 'Nearly', 'half', 'a', 'century', 'later', 'the', 'discovery', 'of', 'the', 'predicted', 'particle', 'brought', 'Higgs', 'who', 'has', 'died', 'aged', '94', 'a', 'share', 'of', 'the', '2013', 'Nobel', 'Prize', 'in', 'Physics', '\n', 'The', 'mechanism', 'Higgs', 'described', 'became', 'a', 'key', 'component', 'of', 'the', 'standard', 'model', 'of', 'particle', 'physics', 'during', 'the', '1970s', 'but', 'the', 'associated', 'particle', 'remained', 'stubbornly', 'elusive', 'Then', 'in', '2012', 'two', 'giant', 'experiments', 'run', 'by', 'more', 'than', '6,000', 'physicists', 'at', 'CERN', 'Europe', 's', 'high', 'energy', 'physics', 'laboratory', 'near', 'Geneva', 'Switzerland', 'discovered', 'something', 'with', 'the', 'appropriate', 'properties', 'By', 'then', 'the', 'particle', 'had', 'achieved', 'fame', 'as', 'the', 'Higgs', 'boson', 'although', 'the', 'self', 'effacing', 'Higgs', 'would', 'usually', 'refer', 'to', 'it', 'as', 'the', 'scalar', 'boson', 'in', 'reference', 'to', 'its', 'key', 'characteristic', 'of', 'having', 'no', 'intrinsic', 'spin', '\n', 'Higgs', 'was', 'born', 'in', 'Newcastle', 'upon', 'Tyne', 'in', '1929', 'but', 'his', 'father', 's', 'work', 'as', 'a', 'sound', 'engineer', 'for', 'the', 'BBC', 'took', 'the', 'family', 'to', 'Bristol', 'where', 'Higgs', 'attended', 'Cotham', 'Grammar', 'School', 'There', 'he', 'spotted', 'several', 'mentions', 'on', 'the', 'honours', 'boards', 'of', 'a', 'previous', 'pupil', 'Paul', 'Dirac', 'who', 'had', 'earned', 'a', 'share', 'of', 'the', '1933', 'Nobel', 'Prize', 'in', 'Physics', 'for', 'his', 'work', 'in', 'quantum', 'mechanics', 'Inspired', 'Higgs', 'took', 'up', 'physics', 'at', 'King', 's', 'College', 'London', 'where', 'he', 'obtained', 'a', 'PhD', 'in', '1954', '\n', 'As', 'a', 'hitch', 'hiking', 'student', 'Higgs', 'had', 'discovered', 'a', 'liking', 'for', 'Edinburgh', 'so', 'in', '1960', 'he', 'was', 'happy', 'to', 'be', 'appointed', 'as', 'a', 'lecturer', 'there', 'He', 'picked', 'up', 'on', 'a', 'long', 'standing', 'interest', 'in', 'symmetry', 'in', 'subatomic', 'particle', 'physics', 'inspired', 'in', 'particular', 'by', 'the', 'work', 'of', 'the', 'future', 'Nobel', 'prizewinner', 'Yoichiro', 'Nambu', 'a', 'Japanese', 'American', 'physicist', 'then', 'at', 'the', 'University', 'of', 'Chicago', 'in', 'Illinois', 'In', 'physics', 'symmetry', 'is', 'linked', 'to', 'the', 'conservation', 'of', 'quantities', 'such', 'as', 'energy', 'momentum', 'and', 'electric', 'charge', 'Working', 'on', 'a', 'theory', 'that', 'had', 'an', 'underlying', 'symmetry', 'but', 'in', 'which', 'particles', 'had', 'no', 'mass', 'Nambu', 'was', 'attempting', 'to', 'generate', 'mass', 'through', 'a', 'mechanism', 'known', 'as', 'spontaneous', 'symmetry', 'breaking', 'However', 'such', 'symmetry', 'breaking', 'would', 'also', 'produce', 'massless', 'particles', 'with', 'zero', 'spin', 'for', 'which', 'there', 'was', 'no', 'evidence', '\n', 'This', 'seemed', 'a', 'dead', 'end', 'but', 'in', '1964', 'Higgs', 'realized', 'that', 'it', 'was', 'possible', 'to', 'get', 'round', 'the', 'difficulty', 'using', 'gauge', 'theory', 'which', 'has', 'the', 'kind', 'of', 'symmetry', 'found', 'for', 'example', 'in', 'the', 'established', 'theory', 'of', 'electromagnetism', 'Higgs', 'showed', 'that', 'the', 'massless', 'particles', 'associated', 'with', 'spontaneous', 'symmetry', 'breaking', 'become', 'absorbed', 'into', 'massive', 'particles', 'He', 'published', 'two', 'short', 'papers', 'on', 'this', 'theme1,2', 'the', 'second', 'of', 'which', 'explicitly', 'predicted', 'a', 'massive', 'spin', 'zero', 'particle', 'It', 'fell', 'to', 'other', 'physicists', 'to', 'realize', 'that', 'this', 'mechanism', 'for', 'spontaneous', 'symmetry', 'breaking', 'was', 'key', 'to', 'a', 'mathematically', 'coherent', 'gauge', 'theory', 'of', 'particle', 'physics', 'that', 'unites', 'the', 'electromagnetic', 'interactions', 'between', 'particles', 'with', 'the', 'weak', 'interaction', 'involved', 'in', 'certain', 'forms', 'of', 'radioactivity', 'This', 'Nobel', 'prizewinning', 'electro', 'weak', 'theory', 'became', 'one', 'of', 'the', 'twin', 'pillars', 'of', 'the', 'standard', 'model', 'of', 'particle', 'physics', 'and', 'was', 'well', 'established', 'by', 'experiments', 'by', 'the', '1990s', '\n', 'The', 'field', 'had', 'seemed', 'something', 'of', 'a', 'scientific', 'backwater', 'in', 'the', '1960s', 'but', 'Higgs', 'was', 'not', 'working', 'alone', 'Two', 'other', 'papers', 'were', 'published', 'in', '1964', 'on', 'the', 'mechanism3,4', 'one', 'appearing', 'just']

t', 'before', 'his', 'own', 'But', 'only', 'Higgs', 'drew', 'attention', 'to', 'the', 'associated', 'massive', 'spin-ze ro', 'particle', 'which', 'in', 'the', '1970s', 'began', 'to', 'be', 'called', 'the', 'Higgs', 'boson', 'The', 'catch y', 'name', 'stuck', '\n', 'Around', 'this', 'time', 'after', 'his', 'marriage', 'broke', 'down', 'Higgs', 'found', 'he', 'was', 'losing', 'his', 'way', 'in', 'theoretical', 'particle', 'physics', 'In', 'addition', 'to', 'teaching', 'he', 'became', 'involved', 'in', 'the', 'union', 'side', 'of', 'university', 'life', 'and', 'wrote', 'few', 'physics', 'papers', 'Nevertheless', 'as', 'appreciation', 'of', 'his', 'work', 'on', 'spontaneous', 'symmetry', 'breaking', 'green', 'he', 'was', 'increasingly', 'asked', 'to', 'give', 'talks', 'a', 'popular', 'title', 'being', 'My', 'life', 'as', 'a', 'boson', 'He', 'retired', 'in', '1996', '\n', 'The', 'award', 'of', 'the', 'Nobel', 'prize', 'to', 'Higgs', 'and', 'Belgian', 'physicist', 'Fran ois', 'Englert', 'the', 'surviving', 'author', 'of', 'the', 'paper', 'published', 'just', 'before', 'Higgs', 's', 'in', '1964', 'came', 'as', 'little', 'surprise', 'to', 'anyone', 'including', 'Higgs', 'because', 'it', 'had', 'been', 'mooted', 'since', 'the', '1980s', 'However', 'interest', 'in', 'the', 'Higgs', 'boson', 'had', 'increased', 'over', 'the', 'years', 'not', 'only', 'among', 'particle', 'physicists', 'but', 'also', 'among', 'the', 'general', 'public', 'and', 'the', 'media', 'It', 'reached', 'fever', 'pitch', 'after', 'the', 'construction', 'of', 'CERN', 's', 'Large', 'Hadron', 'Collider', 'which', 'was', 'billed', 'by', 'many', 'as', 'the', 'machine', 'that', 'would', 'discover', 'the', 'last', 'missing', 'piece', 'of', 'the', 'standard', 'model', 'This', 'intense', 'interest', 'brought', 'fame', 'at', 'a', 'level', 'no', 'one', 'could', 'have', 'imagined', 'in', 'the', '1960s', 'and', 'the', 'quiet', 'physicist', 'became', 'a', 'media', 'star', 'before', 'retiring', 'for', 'a', 'second', 'time', 'when', 'he', 'reached', '85', '\n', 'A', 'keen', 'music', 'lover', 'Higgs', 'had', 'little', 'interest', 'in', 'the', 'trappings', 'of', 'modern', 'technology', 'he', 'famously', 'had', 'no', 'television', 'and', 'did', 'not', 'use', 'the', 'Internet', 'Yet', 'he', 'was', 'far', 'from', 'being', 'remote', 'from', 'the', 'world', 'He', 'had', 'a', 'strong', 'social', 'conscience', 'and', 'was', 'a', 'member', 'of', 'the', 'Campaign', 'for', 'Nuclear', 'Disarmament', 'and', 'Green peace', 'at', 'various', 'times', 'Humble', 'in', 'many', 'respects', 'with', 'an', 'infectious', 'sense', 'of', 'humour', 'he', 'was', 'proud', 'of', 'the', 'work', 'that', 'he', 'had', 'always', 'known', 'was', 'important', 'and', 'which', 'ultimately', 'brought', 'him', 'fame']

In []: # Find the most common words
from collections import Counter

```
# Create a Counter object  
word_freq = Counter(words)  
# Get the most common words  
most_common_words = word_freq.most_common()  
print(most_common_words)
```

[('the', 60), ('of', 34), ('a', 31), ('in', 29), ('Higgs', 20), ('to', 19), ('was', 17), ('had', 13), ('particle', 12), ('he', 11), ('as', 10), ('and', 10), ('for', 9), ('that', 9), ('physics', 9), ('symmetry', 9), ('his', 8), ('\n', 8), ('particles', 7), ('but', 7), ('which', 7), ('at', 6), ('theory', 6), ('no', 6), ('on', 6), ('it', 5), ('The', 5), ('No bel', 5), ('by', 5), ('s', 5), ('with', 5), ('boson', 5), ('work', 5), ('breaking', 5), ('1964', 4), ('physicist', 4), ('papers', 4), ('mechanism', 4), ('mass', 4), ('massive', 4), ('became', 4), ('He', 4), ('interest', 4), ('spontaneou s', 4), ('two', 3), ('second', 3), ('predicted', 3), ('brought', 3), ('key', 3), ('standard', 3), ('model', 3), ('associated', 3), ('physicists', 3), ('fame', 3), ('would', 3), ('spin', 3), ('This', 3), ('published', 3), ('this', 3), ('on e', 3), ('not', 3), ('before', 3), ('few', 2), ('theoretical', 2), ('University', 2), ('Edinburgh', 2), ('wrote', 2), ('short', 2), ('could', 2), ('give', 2), ('mathematically', 2), ('ultimately', 2), ('because', 2), ('described', 2), ('paper', 2), ('drew', 2), ('attention', 2), ('who', 2), ('has', 2), ('share', 2), ('Prize', 2), ('Physics', 2), ('1970 s', 2), ('experiments', 2), ('CERN', 2), ('energy', 2), ('discovered', 2), ('something', 2), ('then', 2), ('took', 2), ('where', 2), ('up', 2), ('be', 2), ('there', 2), ('Nambu', 2), ('In', 2), ('such', 2), ('an', 2), ('known', 2), ('Howe ver', 2), ('also', 2), ('massless', 2), ('zero', 2), ('seemed', 2), ('gauge', 2), ('found', 2), ('established', 2), ('It', 2), ('other', 2), ('interactions', 2), ('weak', 2), ('involved', 2), ('1960s', 2), ('just', 2), ('only', 2), ('time', 2), ('after', 2), ('life', 2), ('being', 2), ('little', 2), ('among', 2), ('media', 2), ('reached', 2), ('many', 2), ('from', 2), ('During', 1), ('weeks', 1), ('summer', 1), ('Peter', 1), ('UK', 1), ('outlining', 1), ('ideas', 1), ('fundamental', 1), ('building', 1), ('blocks', 1), ('Universe', 1), ('His', 1), ('aim', 1), ('rescue', 1), ('appealing', 1), ('unrealistic', 1), ('measurable', 1), ('consequence', 1), ('proposal', 1), ('existence', 1), ('new', 1), ('Nearly', 1), ('half', 1), ('century', 1), ('later', 1), ('discovery', 1), ('died', 1), ('aged', 1), ('94', 1), ('2013', 1), ('component', 1), ('during', 1), ('remained', 1), ('stubbornly', 1), ('elusive', 1), ('Then', 1), ('2012', 1), ('giant', 1), ('run', 1), ('more', 1), ('than', 1), ('6,000', 1), ('Europe', 1), ('high', 1), ('laboratory', 1), ('near', 1), ('Geneva', 1), ('Switzerland', 1), ('appropriate', 1), ('properties', 1), ('By', 1), ('achieved', 1), ('although', 1), ('self', 1), ('effacing', 1), ('usually', 1), ('refer', 1), ('scalar', 1), ('reference', 1), ('its', 1), ('characteristic', 1), ('having', 1), ('intrinsic', 1), ('born', 1), ('Newcastle', 1), ('upon', 1), ('Tyne', 1), ('1929', 1), ('father', 1), ('sound', 1), ('engineer', 1), ('BBC', 1), ('family', 1), ('Bristol', 1), ('attended', 1), ('Coatham', 1), ('Grammar', 1), ('School', 1), ('There', 1), ('spotted', 1), ('several', 1), ('mentions', 1), ('honours', 1), ('boards', 1), ('previous', 1), ('pupil', 1), ('Paul', 1), ('Dirac', 1), ('earned', 1), ('1933', 1), ('quantum', 1), ('mechanics', 1), ('Inspired', 1), ('King', 1), ('College', 1), ('London', 1), ('obtained', 1), ('PhD', 1), ('1954', 1), ('As', 1), ('hitch', 1), ('hiking', 1), ('student', 1), ('liking', 1), ('so', 1), ('1960', 1), ('happy', 1), ('appointed', 1), ('lecturer', 1), ('picked', 1), ('long', 1), ('standing', 1), ('subatomic', 1), ('inspired', 1), ('particular', 1), ('future', 1), ('prizewinner', 1), ('Yoichiro', 1), ('Japanese', 1), ('American', 1), ('Chicago', 1), ('Illinois', 1), ('is', 1), ('linked', 1), ('conservation', 1), ('quantities', 1), ('momentum', 1), ('electric', 1), ('charge', 1), ('Working', 1), ('underlying', 1), ('attempting', 1), ('generate', 1), ('through', 1), ('produce', 1), ('evidence', 1), ('dead', 1), ('end', 1), ('realized', 1), ('possible', 1), ('get', 1), ('round', 1), ('difficulty', 1), ('using', 1), ('kind', 1), ('example', 1), ('electromagnetism', 1), ('showed', 1), ('become', 1), ('absorbed', 1), ('into', 1), ('theme1', 2), ('explicitly', 1), ('fell', 1), ('realize', 1), ('coherent', 1), ('unites', 1), ('electromagnetic', 1), ('between', 1), ('certain', 1), ('forms', 1), ('radioactivity', 1), ('prizewinning', 1), ('electro', 1), ('twin', 1), ('pillars', 1), ('well', 1), ('1990s', 1), ('field', 1), ('scientific', 1), ('backwater', 1), ('working', 1), ('alone', 1), ('Two', 1), ('were', 1), ('mechanism3,4', 1), ('appearing', 1), ('own', 1), ('But', 1), ('spin-zero', 1), ('began', 1), ('called', 1), ('catchy', 1), ('name', 1), ('stuck', 1), ('Around', 1), ('marriage', 1), ('broke', 1), ('down', 1), ('losing', 1), ('way', 1), ('addition', 1), ('teaching', 1), ('union', 1), ('side', 1), ('university', 1), ('Nevertheless', 1), ('appreciation', 1), ('grew', 1), ('increasingly', 1), ('asked', 1), ('talks', 1), ('popular', 1), ('title', 1), ('My', 1), ('retired', 1), ('1996', 1), ('award', 1), ('prize', 1), ('Belgian', 1), ('François', 1), ('Englert', 1), ('surviving', 1), ('author', 1), ('came', 1), ('surprise', 1), ('anyone', 1), ('including', 1), ('been', 1), ('moot ed', 1), ('since', 1), ('1980s', 1), ('increased', 1), ('over', 1), ('years', 1), ('general', 1), ('public', 1), ('feve

```
r', 1), ('pitch', 1), ('construction', 1), ('Large', 1), ('Hadron', 1), ('Collider', 1), ('billed', 1), ('machine', 1), ('discover', 1), ('last', 1), ('missing', 1), ('piece', 1), ('intense', 1), ('level', 1), ('have', 1), ('imagined', 1), ('quiet', 1), ('star', 1), ('retiring', 1), ('when', 1), ('85', 1), ('A', 1), ('keen', 1), ('music', 1), ('lover', 1), ('trappings', 1), ('modern', 1), ('technology', 1), ('famously', 1), ('television', 1), ('did', 1), ('use', 1), ('Internet', 1), ('Yet', 1), ('far', 1), ('remote', 1), ('world', 1), ('strong', 1), ('social', 1), ('conscience', 1), ('member', 1), ('Campaign', 1), ('Nuclear', 1), ('Disarmament', 1), ('Greenpeace', 1), ('various', 1), ('times', 1), ('Humble', 1), ('respects', 1), ('infectious', 1), ('sense', 1), ('humour', 1), ('proud', 1), ('always', 1), ('important', 1), ('him', 1)]
```

Most common words under a part of speech

```
In [ ]: # Create a Counter object
from collections import defaultdict
```

```
In [ ]: count_pos = defaultdict(Counter)
for token in doc:
    count_pos[token.pos_][token.text] += 1

for pos in count_pos:
    print(pos)
    #for word, count in count_pos[pos].most_common():
    #    print(f"{word}: {count}")
    print(count_pos[pos].most_common())
    print("")
```

ADP

[('of', 34), ('in', 29), ('to', 12), ('for', 9), ('as', 8), ('at', 6), ('on', 6), ('by', 5), ('with', 5), ('up', 2), ('In', 2), ('before', 2), ('among', 2), ('from', 2), ('During', 1), ('during', 1), ('than', 1), ('near', 1), ('By', 1), ('As', 1), ('through', 1), ('round', 1), ('into', 1), ('between', 1), ('Around', 1), ('down', 1), ('over', 1), ('after', 1)]

DET

[('the', 60), ('a', 31), ('no', 6), ('The', 5), ('this', 3), ('an', 2), ('This', 2), ('half', 1), ('A', 1)]

ADJ

[('massive', 4), ('spontaneous', 4), ('key', 3), ('standard', 3), ('few', 2), ('theoretical', 2), ('short', 2), ('second', 2), ('associated', 2), ('such', 2), ('massless', 2), ('other', 2), ('weak', 2), ('little', 2), ('many', 2), ('fundamental', 1), ('appealing', 1), ('unrealistic', 1), ('measurable', 1), ('new', 1), ('aged', 1), ('elusive', 1), ('giant', 1), ('more', 1), ('high', 1), ('appropriate', 1), ('scalar', 1), ('intrinsic', 1), ('sound', 1), ('several', 1), ('previous', 1), ('quantum', 1), ('happy', 1), ('subatomic', 1), ('particular', 1), ('future', 1), ('Japanese', 1), ('American', 1), ('electric', 1), ('dead', 1), ('possible', 1), ('coherent', 1), ('electromagnetic', 1), ('certain', 1), ('twin', 1), ('scientific', 1), ('own', 1), ('catchy', 1), ('involved', 1), ('popular', 1), ('Belgian', 1), ('general', 1), ('Large', 1), ('last', 1), ('missing', 1), ('intense', 1), ('quiet', 1), ('keen', 1), ('modern', 1), ('remote', 1), ('strong', 1), ('social', 1), ('various', 1), ('infectious', 1), ('proud', 1), ('important', 1)]

NOUN

[('particle', 12), ('physics', 9), ('symmetry', 9), ('particles', 7), ('theory', 6), ('boson', 5), ('work', 5), ('breaking', 5), ('physicist', 4), ('papers', 4), ('mechanism', 4), ('mass', 4), ('interest', 4), ('model', 3), ('physicists', 3), ('fame', 3), ('spin', 3), ('paper', 2), ('attention', 2), ('share', 2), ('1970s', 2), ('experiments', 2), ('energy', 2), ('gauge', 2), ('interactions', 2), ('1960s', 2), ('time', 2), ('life', 2), ('media', 2), ('weeks', 1), ('summer', 1), ('ideas', 1), ('building', 1), ('blocks', 1), ('aim', 1), ('consequence', 1), ('proposal', 1), ('existence', 1), ('century', 1), ('discovery', 1), ('component', 1), ('laboratory', 1), ('properties', 1), ('self', 1), ('effacing', 1), ('reference', 1), ('characteristic', 1), ('father', 1), ('engineer', 1), ('family', 1), ('mentions', 1), ('honours', 1), ('boards', 1), ('pupil', 1), ('mechanics', 1), ('PhD', 1), ('hitch', 1), ('hiking', 1), ('student', 1), ('liking', 1), ('lecturer', 1), ('prizewinner', 1), ('conservation', 1), ('quantities', 1), ('momentum', 1), ('charge', 1), ('evidence', 1), ('end', 1), ('difficulty', 1), ('kind', 1), ('example', 1), ('electromagnetism', 1), ('second', 1), ('forms', 1), ('radioactivity', 1), ('electro', 1), ('pillars', 1), ('field', 1), ('backwater', 1), ('name', 1), ('marriage', 1), ('way', 1), ('addition', 1), ('teaching', 1), ('union', 1), ('side', 1), ('university', 1), ('appreciation', 1), ('talks', 1), ('title', 1), ('award', 1), ('prize', 1), ('author', 1), ('surprise', 1), ('years', 1), ('public', 1), ('fever', 1), ('pitch', 1), ('construction', 1), ('machine', 1), ('piece', 1), ('level', 1), ('star', 1), ('music', 1), ('lover', 1), ('trappings', 1), ('technology', 1), ('television', 1), ('Internet', 1), ('world', 1), ('conscience', 1), ('member', 1), ('times', 1), ('respects', 1), ('sense', 1), ('humour', 1)]

NUM

[('1964', 4), ('two', 3), ('zero', 2), ('one', 2), ('94', 1), ('2013', 1), ('2012', 1), ('6,000', 1), ('1929', 1), ('1933', 1), ('1954', 1), ('1960', 1), ('1990s', 1), ('Two', 1), ('spin-zero', 1), ('1996', 1), ('1980s', 1), ('85', 1)]

PUNCT

[(' ', 58), ('.', 36), ('-', 7), ('-', 4), (''', 4), ('''', 4)]

PROPN

[('Higgs', 20), ('Nobel', 5), ('University', 2), ('Edinburgh', 2), ('Prize', 2), ('Physics', 2), ('CERN', 2), ('Nambu', 2), ('Peter', 1), ('UK', 1), ('Universe', 1), ('Europe', 1), ('Geneva', 1), ('Switzerland', 1), ('Newcastle', 1), ('Tyn e', 1), ('BBC', 1), ('Bristol', 1), ('Cotham', 1), ('Grammar', 1), ('School', 1), ('Paul', 1), ('Dirac', 1), ('King', 1), ('College', 1), ('London', 1), ('Yoichiro', 1), ('Chicago', 1), ('Illinois', 1), ('theme1,2', 1), ('mechanism3,4', 1), ('François', 1), ('Englert', 1), ('Hadron', 1), ('Collider', 1), ('Campaign', 1), ('Nuclear', 1), ('Disarmament', 1), ('Greenpeace', 1)]

VERB

[('had', 6), ('became', 4), ('predicted', 3), ('brought', 3), ('published', 3), ('wrote', 2), ('give', 2), ('described', 2), ('drew', 2), ('discovered', 2), ('took', 2), ('known', 2), ('seemed', 2), ('found', 2), ('established', 2), ('reached', 2), ('outlining', 1), ('rescue', 1), ('died', 1), ('remained', 1), ('run', 1), ('achieved', 1), ('refer', 1), ('having', 1), ('born', 1), ('attended', 1), ('spotted', 1), ('earned', 1), ('Inspired', 1), ('obtained', 1), ('appointed', 1), ('picked', 1), ('standing', 1), ('inspired', 1), ('linked', 1), ('Working', 1), ('underlying', 1), ('attempting', 1), ('generate', 1), ('produce', 1), ('was', 1), ('realized', 1), ('get', 1), ('using', 1), ('has', 1), ('showed', 1), ('associated', 1), ('become', 1), ('absorbed', 1), ('fell', 1), ('realize', 1), ('unites', 1), ('involved', 1), ('prizewinning', 1), ('working', 1), ('appearing', 1), ('began', 1), ('called', 1), ('stuck', 1), ('broke', 1), ('losing', 1), ('grew', 1), ('asked', 1), ('retired', 1), ('surviving', 1), ('came', 1), ('including', 1), ('mooted', 1), ('increased', 1), ('billed', 1), ('discover', 1), ('imagined', 1), ('retiring', 1), ('use', 1), ('Humble', 1)]

PRON

[('he', 11), ('his', 8), ('which', 7), ('that', 6), ('it', 5), ('He', 4), ('who', 2), ('something', 2), ('It', 2), ('His', 1), ('its', 1), ('there', 1), ('This', 1), ('My', 1), ('anyone', 1), ('one', 1), ('him', 1)]

AUX

[('was', 16), ('had', 7), ('would', 3), ('could', 2), ('be', 2), ('being', 2), ('has', 1), ('is', 1), ('were', 1), ('been', 1), ('have', 1), ('did', 1)]

PART

[('to', 7), ('s', 5), ('not', 3)]

ADV

[('mathematically', 2), ('ultimately', 2), ('then', 2), ('However', 2), ('also', 2), ('just', 2), ('only', 2), ('Nearly', 1), ('later', 1), ('stubbornly', 1), ('Then', 1), ('usually', 1), ('There', 1), ('so', 1), ('there', 1), ('long', 1), ('explicitly', 1), ('well', 1), ('alone', 1), ('Nevertheless', 1), ('increasingly', 1), ('as', 1), ('famously', 1), ('far', 1), ('always', 1)]

CCONJ

[('and', 10), ('but', 7), ('But', 1), ('Yet', 1)]

SCONJ

[('that', 3), ('because', 2), ('where', 2), ('although', 1), ('upon', 1), ('after', 1), ('as', 1), ('before', 1), ('since', 1), ('when', 1)]

SPACE

```
[('\'\n', 8)]
```

Find a subject/object relationship through the dependency parser in any sentence.

```
In [ ]: def pr_tree(word, level):
    if word.is_punct:
        return

    for child in word.lefts:
        pr_tree(child, level+1)
    if word.dep_ == 'nsubj' or word.dep_ == 'dobj':
        print('\t'* level + word.text + ' - ' + word.dep_)
    #print('\t'* level + word.text + ' - ' + word.dep_)
    for child in word.rights:
        pr_tree(child, level+1)
```

```
In [ ]: for sentence in doc.sents:
    pr_tree(sentence.root, 0)
    print('-----')
```

Higgs - nsubj
papers - dobj
 ideas - dobj

 that - nsubj
 mass - dobj

 aim - nsubj
 theory - dobj
 that - nsubj
 particles - nsubj
 it - nsubj
 mass - dobj

 paper - nsubj
 attention - dobj
 it - nsubj
 existence - dobj

 discovery - nsubj
 Higgs - dobj
 who - nsubj
 94 - dobj

 Higgs - nsubj
 particle - nsubj

 experiments - nsubj
 something - dobj

 particle - nsubj
 fame - dobj
 Higgs - nsubj

spin - dobj

 work - nsubj
 family - dobj

 Higgs - nsubj
 School - dobj

 he - nsubj
 mentions - dobj

 who - nsubj
 share - dobj

 Higgs - nsubj

physics - dobj

 he - nsubj
 PhD - dobj

 Higgs - nsubj

 liking - dobj

 he - nsubj

 He - nsubj

 that - nsubj
 symmetry - dobj
 particles - nsubj
 mass - dobj

 Nambu - nsubj

 mass - dobj

 breaking - nsubj

 particles - dobj

 This - nsubj

 Higgs - nsubj

 it - nsubj

 theory - dobj

 which - nsubj
 kind - dobj

 Higgs - nsubj

 particles - nsubj

 He - nsubj

 papers - dobj

 second - nsubj

 particle - dobj

 It - nsubj

 mechanism - nsubj

 that - nsubj
 interactions - dobj

 theory - nsubj

 field - nsubj

 something - dobj

Higgs - nsubj

one - nsubj

Higgs - nsubj

attention - dobj

which - nsubj

name - nsubj

marriage - nsubj

Higgs - nsubj

he - nsubj

way - dobj

he - nsubj

papers - dobj

appreciation - nsubj

talks - dobj

title - nsubj

He - nsubj

award - nsubj

Higgs - nsubj

's - dobj

interest - nsubj

It - nsubj

pitch - dobj

interest - nsubj

fame - dobj

one - nsubj

physicist - nsubj

Higgs - nsubj

interest - dobj

he - nsubj

that - nsubj

piece - dobj

he - nsubj

85 - dobj

```
television - dobj
Internet - dobj
-----
he - nsubj
-----
He - nsubj
conscience - dobj
-----
he - nsubj
that - dobj
he - nsubj
which - nsubj
fame - dobj
```

Show the most common Entities and their types.

```
In [ ]: count_ent = defaultdict(Counter)
for entity in doc.ents:
    count_ent[entity.label_][entity.text] += 1

for ent in count_ent:
    print(ent)
#    for word, count in count_pos[pos].most_common():
#        print(f'{word}: {count}')
    print(count_ent[ent].most_common())
    print("")
```

DATE

```
[('1964', 3), ('the 1970s', 2), ('the 1960s', 2), ('a few weeks', 1), ('the summer of 1964', 1), ('Nearly half a century later', 1), ('aged 94', 1), ('2013', 1), ('2012', 1), ('1929', 1), ('1933', 1), ('1954', 1), ('1960', 1), ('the 1990s', 1), ('1996', 1), ('the 1980s', 1), ('the years', 1)]
```

PERSON

```
[('Higgs', 18), ('Peter Higgs', 1), ('Paul Dirac', 1), ('Yoichiro Nambu', 1), ('Fran ois Englert', 1)]
```

ORG

```
[('CERN', 2), ('the University of Edinburgh', 1), ('Universe', 1), ('BBC', 1), ('Bristol', 1), ('Cotham Grammar School', 1), ('King's College London', 1), ('the University of Chicago', 1), ('the Campaign for Nuclear Disarmament and Greenpeace', 1)]
```

GPE

```
[('UK', 1), ('Geneva', 1), ('Switzerland', 1), ('Newcastle', 1), ('Tyne', 1), ('Edinburgh', 1), ('Illinois', 1), ('Nambu', 1)]
```

CARDINAL

```
[('two', 3), ('zero', 2), ('more than 6,000', 1), ('Two', 1), ('one', 1), ('spin-zero', 1), ('85', 1)]
```

ORDINAL

```
[('second', 3)]
```

WORK_OF_ART

```
[('Nobel Prize in Physics', 2), ('Nobel', 2), ('PhD', 1), ('the Nobel prize to Higgs', 1)]
```

LOC

```
[('Europe', 1)]
```

NORP

```
[('Japanese American', 1), ('Belgian', 1)]
```

Find Entities and their dependency (hint: entity.root.head)

```
In [ ]: # entity.root.head did not appear anywhere in spaCy documentation!
# Print the entities and their dependencies
for entity in doc.ents:
    print(entity.text, entity.label_, entity.root.dep_)
```

a few weeks DATE pobj
the summer of 1964 DATE pobj
Peter Higgs PERSON nsubj
the University of Edinburgh ORG pobj
UK GPE appos
two CARDINAL nummod
Universe ORG pobj
second ORDINAL amod
Nearly half a century later DATE advmod
Higgs PERSON dobj
aged 94 DATE advcl
2013 DATE nummod
Nobel Prize in Physics WORK_OF_ART pobj
Higgs PERSON nsubj
the 1970s DATE pobj
2012 DATE pobj
two CARDINAL nummod
more than 6,000 CARDINAL nummod
CERN ORG pobj
Europe LOC poss
Geneva GPE pobj
Switzerland GPE appos
Higgs PERSON compound
Higgs PERSON nsubj
Higgs PERSON nsubjpass
Newcastle GPE pobj
Tyne GPE pobj
1929 DATE pobj
BBC ORG pobj
Bristol ORG pobj
Higgs PERSON nsubj
Cotham Grammar School ORG dobj
Paul Dirac PERSON appos
1933 DATE nummod
Nobel Prize in Physics WORK_OF_ART pobj
Higgs PERSON nsubj
King's College London ORG pobj
PhD WORK_OF_ART dobj
1954 DATE pobj
Higgs PERSON nsubj
Edinburgh GPE pobj
1960 DATE pobj
Nobel WORK_OF_ART compound
Yoichiro Nambu PERSON appos
Japanese American NORG amod

the University of Chicago ORG pobj
Illinois GPE pobj
Nambu GPE nsubj
zero CARDINAL nummod
1964 DATE pobj
Higgs PERSON nsubj
Higgs PERSON nsubj
two CARDINAL nummod
second ORDINAL nsubj
zero CARDINAL nummod
Nobel WORK_OF_ART compound
the 1990s DATE pobj
the 1960s DATE pobj
Higgs PERSON nsubj
Two CARDINAL nummod
1964 DATE pobj
one CARDINAL nsubj
Higgs PERSON nsubj
spin-zero CARDINAL compound
the 1970s DATE pobj
Higgs PERSON compound
Higgs PERSON nsubj
1996 DATE pobj
the Nobel prize to Higgs WORK_OF_ART prep
Belgian NORP amod
François Englert PERSON conj
Higgs PERSON nsubj
1964 DATE pobj
Higgs PERSON pobj
the 1980s DATE pobj
Higgs PERSON compound
the years DATE pobj
CERN ORG poss
the 1960s DATE pobj
second ORDINAL amod
85 CARDINAL dobj
Higgs PERSON nsubj
the Campaign for Nuclear Disarmament and Greenpeace ORG pobj

Find the most similar words in the article

```
In [ ]: # Get the word vectors for each token in the text
word_vectors = [token.vector for token in doc]
```

```
# Calculate the similarity between each pair of words
similarities = []
# Loop through all of the words in the text
for i in range(len(word_vectors)):
    # Loop through all of the words in the text that come after word i
    for j in range(i+1, len(word_vectors)):
        # Calculate the similarity between word vectors i and j
        similarity = word_vectors[i].dot(word_vectors[j])

        # Ignore if the words themselves are the same
        if doc[i].lemma_ != doc[j].lemma_:
            # Append the word pair and their similarity score
            similarities.append((doc[i].text, doc[j].text, similarity))

# Sort the similarities in descending order
similarities.sort(key=lambda x: x[2], reverse=True)
```

```
In [ ]: # Print the top most similar word pairs
for word1, word2, similarity in similarities[:20]:
    print(f"{word1} - {word2}: {similarity}")
```