0.1 SpaCy examples

```
[5]: import spacy
nlp = spacy.load("en_core_web_sm")
doc = nlp("Apple is looking at buying U.K. startup for $11.1 million.")
for token in doc:
    print(token.text, token.lemma_, token.pos_, token.tag_, token.dep_,
          token.shape_, token.is_alpha, token.is_stop)
```

Apple Apple PROPN NNP nsubj Xxxxx True False
is be AUX VBZ aux xx True True
looking look VERB VBG ROOT xxxx True False
at at ADP IN prep xx True True
buying buy VERB VBG pcomp xxxx True False
U.K. U.K. PROPN NNP compound X.X. False False
startup startup NOUN NN dobj xxxx True False
for for ADP IN prep xxx True True
$ $ SYM $ quantmod $ False False
11.1 11.1 NUM CD compound dd.d False False
million million NUM CD pobj xxxx True False
. . PUNCT . punct . False False

0.1.1 Display syntactic dependences

```
[6]: import spacy
    from spacy import displacy
    nlp = spacy.load("en_core_web_sm")
doc = nlp("Microsoft plans to buy Activision for $69 billion.")
displacy.render(doc, style = "dep")

<IPython.core.display.HTML object>
```

```
[7]: import spacy
    from spacy import displacy
    nlp = spacy.load("en_core_web_sm")
doc = nlp("Microsoft plans to buy Activision for $69 billion.")
```
0.1.2 Use directly just the tokenizer component.

```python
from spacy.lang.en import English
nlp = English()
tokenizer = nlp.tokenizer
tokens = tokenizer("U.S. economy is healing, but there's a long way to go. "
                   "The spread of Covid-19 led to surge in orders for factory robots.")
for token in tokens:
    print(token.text, end = ' ')
print()
```

U.S. economy is healing, but there's a long way to go. The spread of Covid-19 led to surge in orders for factory robots.

0.1.3 Use only the tokenizer component by disabling other pipeline modules.

```python
import spacy
nlp = spacy.load("en_core_web_sm", exclude = ['tagger', 'ner', 'parser'])
doc = nlp("U.S. economy is healing, but there's a long way to go. "
          "The spread of Covid-19 led to surge in orders for factory robots.")
for token in doc:
    print(token.text, end = ' ')
print()
```

U.S. economy is healing, but there's a long way to go. The spread of Covid-19 led to surge in orders for factory robots.

0.1.4 Use a special sentencizer that does not require syntactic parsing, for efficiency.

```python
from spacy.lang.en import English
nlp = English()
nlp.add_pipe("sentencizer")
doc = nlp("U.S. economy is healing, but there's a long way to go. "
          "The spread of Covid-19 led to surge in orders for factory robots.")
# Print tokens, one sentence per line.
```
for sent in doc.sents:
    for token in sent:
        print (token, end = ' ')
    print()

U.S. economy is healing, but there 's a long way to go. The spread of Covid-19 led to surge in orders for factory robots.

0.1.5 By default, it seems spaCy does not allow sentences to span newline characters.

```
[11]: nlp = spacy.load("en_core_web_sm")

stanza = "I am a contract-drafting em, \n      "The loyalest of lawyers! \n      "I draw up terms for deals 'twixt firms \n      "To service my employers!"
print(stanza, '\n')

doc = nlp(stanza)

# Print tokens, one sentence per line.
for sent in doc.sents:
    for token in sent:
        print (token, end = ' ')
    print()
```

I am a contract-drafting em, 
The loyalest of lawyers!  
I draw up terms for deals 'twixt firms  
To service my employers!

I am a contract - drafting em ,

The loyalest of lawyers !

I draw up terms for deals ' twixt firms

To service my employers !

0.1.6 Replace newlines with white spaces to enable better sentence segmentation.

```
[12]: stanza = "I am a contract-drafting em, " 
     "The loyalest of lawyers! " 
     "I draw up terms for deals 'twixt firms " 
     "To service my employers!"
print(stanza)
print()
```

3
I am a contract-drafting em, The loyalest of lawyers! I draw up terms for deals 'twixt firms To service my employers!

I am a contract-drafting em, The loyalest of lawyers!
I draw up terms for deals 'twixt firms To service my employers!

[ ]: