

Installations, if needed

```
In [ ]: !pip install spacy==2.0.12 # Above 2.0.12 doesn't seem work with the neuralcoref r
        resolution (at least 2.0.13 and 2.0.16 don't)
```

Importing Libraries

```
In [2]: import spacy
        from spacy import displacy
        from collections import Counter
        import re
        import os
        import pandas as pd
        import sys
```

Loading and previewing our export from OntoNotes5

```
In [3]: JSON_FILENAME = 'ner_output_1.json'
        FILEPATH_TO_JSON = "onto_sql_output/"

        onto_import = pd.read_json(FILEPATH_TO_JSON + JSON_FILENAME)
```

```
In [4]: onto_import.head()
```

Out[4]:

	document_text	ner_end_word_index	ner_start_word_index	ner_string	ner_type	onto_sentence_r
0	The explosion in Yemen did not help an already...	32	32	Israeli	NORP	(TOP (S (S (NP-SE (NP (DT The) (NN explosion))...
1	The explosion in Yemen did not help an already...	34	34	Palestinian	NORP	(TOP (S (S (NP-SE (NP (DT The) (NN explosion))...
2	The explosion in Yemen did not help an already...	3	3	Yemen	GPE	(TOP (S (S (NP-SE (NP (DT The) (NN explosion))...
3	The explosion in Yemen did not help an already...	37	37	Egypt	GPE	(TOP (S (S (NP-SE (NP (DT The) (NN explosion))...
4	The explosion in Yemen did not help an already...	42	39	the next several days	DATE	(TOP (S (S (NP-SE (NP (DT The) (NN explosion))...

Downloading and loading the large spaCy English pipeline

```
In [5]: # Download the english medium-sized pipeline
! python -m spacy download en_core_web_lg
```

```
Collecting en_core_web_lg==2.0.0 from https://github.com/explosion/spacy-models/releases/download/en_core_web_lg-2.0.0/en_core_web_lg-2.0.0.tar.gz#egg=en_core_web_lg==2.0.0
```

```
Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web_lg-2.0.0/en_core_web_lg-2.0.0.tar.gz (852.3MB)
```

[illegible]

```
Installing collected packages: en-core-web-lg
```

Running setup.py install for en-core-web-lg ... done

Successfully installed en-core-web-lg-2.0.0

Linking successful

```
/anaconda/envs/ear/lib/python3.6/site-packages/en_core_web_lg -->
```

```
/anaconda/envs/ear/lib/python3.6/site-packages/spacy/data/en_core_web_lg
```

You can now load the model via `spacy.load('en_core_web_lg')`

```
In [7]: nlp = spacy.load('en_core_web_lg')
```

```
# Testing one sentence
```

```
doc = nlp(onto_import.loc[1, 'sentence_string'])
```

Looping through dependency parsing for all sentences

```
In [12]: onto_import["spacy_parse"] = onto_import.apply(lambda x: nlp(x["sentence_string"])
, axis=1)
```

```
In [15]: onto_import.head()
```

Out[15]:

	document_text	ner_end_word_index	ner_start_word_index	ner_string	ner_type	onto_sentence_r
0	The explosion in Yemen did not help an already...	32	32	Israeli	NORP	(TOP (S (S (NP-SE (NP (DT The) (NN explosion))...
1	The explosion in Yemen did not help an already...	34	34	Palestinian	NORP	(TOP (S (S (NP-SE (NP (DT The) (NN explosion))...
2	The explosion in Yemen did not help an already...	3	3	Yemen	GPE	(TOP (S (S (NP-SE (NP (DT The) (NN explosion))...
3	The explosion in Yemen did not help an already...	37	37	Egypt	GPE	(TOP (S (S (NP-SE (NP (DT The) (NN explosion))...
4	The explosion in Yemen did not help an already...	42	39	the next several days	DATE	(TOP (S (S (NP-SE (NP (DT The) (NN explosion))...

Viewing the text, with highlighted named entities

```
In [35]: # Choose an entry integer to see its text and the parse below.
ENTRY = 47

displacy.render(onto_import.loc[ENTRY,"spacy_parse"], jupyter=True, style='ent')
```

First **ORDINAL** they brought in other Hasidic Jews **NORP** , then **Mexicans** **NORP** , **Palestinians** **NORP** , **Ukrainians** **NORP** .

Viewing dependencies

```
In [36]: displacy.render(onto_import.loc[ENTRY,"spacy_parse"], jupyter=True, style='dep')
```

