

NeuralCoref: Coreference Resolution in spaCy

Dana Nica 341C4

1 Introducere

În lingvistică, coreferința, apare atunci când 2 sau mai multe expresii dintr-un text referă aceeași persoană sau același obiect, cu alte cuvinte au același referent. Rezoluția coreferințelor este procesul de asociere a mențiunilor care referă aceeași entitate.

Exemplu: "When Emily PROPER is at the front door, she PRONOMINAL must be leaving the house"

Substantivul propriu *Emily* și pronumele *she* referă aceeași persoană- Emily.

Tipuri de coreferințe

- **Anaphora** (anaphora presupune folosirea unei expresii care depinde de o expresie specifică anterioară)

"Because David PROPER was so cold, he PRONOMINAL put on his PRONOMINAL coat."

" My sister NOMINAL has a dog NOMINAL . She PRONOMINAL loves him PRONOMINAL ."

- **Cataphora** (folosirea unei expresii sau a unui cuvânt care referă o expresie ulterioară, mai specifică, în discurs)

"Because he PRONOMINAL was so cold, David PROPER put on his PRONOMINAL coat."

" His PRONOMINAL friends have been criticizing Jim PROPER for exaggerating."

- **Split antecedents**

" Carol **PROPER** told Bob **PROPER** to attend the party. They **PRONOMINAL** arrived together."

They referă atât pe Bob cât și pe Carol.

- **Coreferring noun phrases**

" Some of our colleagues **NOMINAL** are going to be supportive. These kinds of people **NOMINAL** will earn our gratitude."

2 NeuralCoref

NeuralCoref[3] este o bibliotecă open source, integrată ca o extensie pipeline în spaCy 2.0 care adnotează și rezolvă coreferințele folosind o rețea neurală. Pentru a folosi NeuralCoref, se încarcă modelul descărcat și se procesează textul cu spaCy. NeuralCoref rezolvă coreferințele și le adnotează ca attribute ale obiectelor spaCy Doc, Span și Token sub dicționarul `...`. Există posibilitatea de a reantrena un model, sau de a extinde la o altă limbă.[5][6]

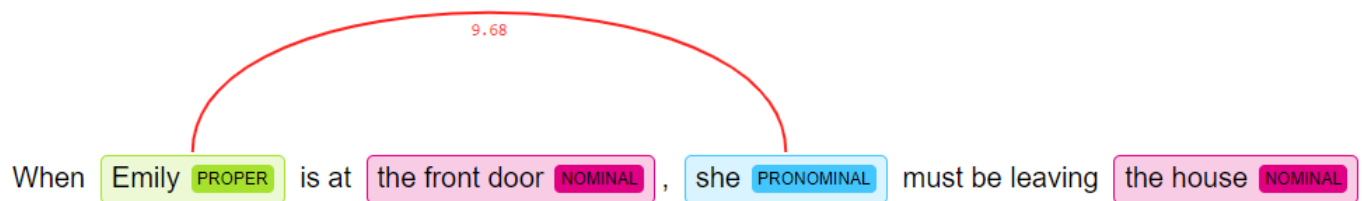


Figura 1: Obținută cu NeuralCoref-Viz[2]

Listing 1: Rezolvarea coreferințelor

```
import spacy
nlp = spacy.load('en_coref_sm')

coref = "When Emily is at the front door, she must be leaving the house"
resolved_coref = nlp(coref).doc._.coref_resolved

print(resolved_coref)
```

When Emily is at the front door, Emily must be leaving the house.

3 Concluzii

NeuralCoref rezolvă doar coreferințele de tipul anaphora. Pentru celelalte cazuri nu identifică complet(Figura 2) sau corect(Figura 3, 4) coreferințele.

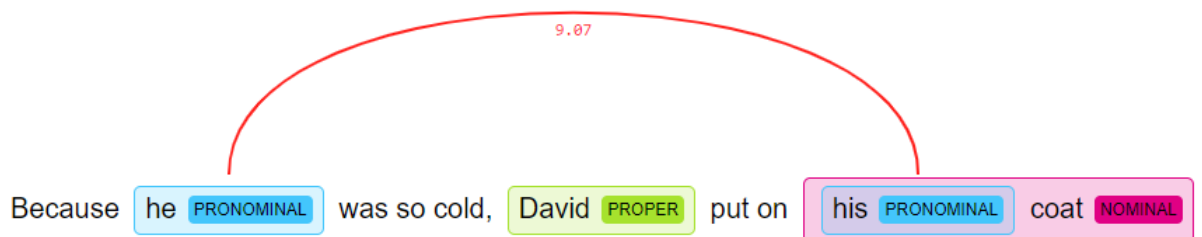


Figura 2: Obținută cu NeuralCoref-Viz[2]

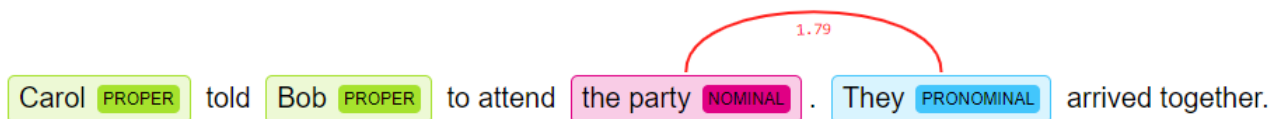


Figura 3: Obținută cu NeuralCoref-Viz[2]

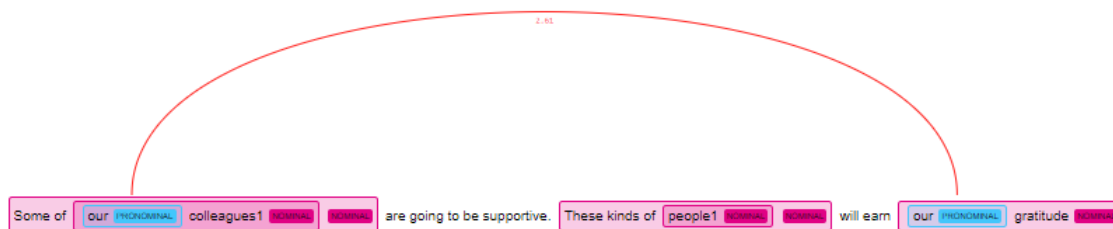


Figura 4: Obținută cu NeuralCoref-Viz[2]

În anexa A sunt prezentate scenarii de utilizare ce conțin coreferințe, iar în anexa B sunt prezentate rezultatele obținute folosind NeuralCoref pe întreaga conversație, informație deținută de agent a fost analizată separat.

Bibliografie

[1] [spaCy 101](#)

[2] NeuralCoref is accompanied by a visualization client [NeuralCoref-Viz](#)

- [3] [NeuralCoref: Coreference Resolution în spaCy with Neural Networks](#)
- [4] [State-of-the-art neural coreference resolution for chatbots](#)
- [5] [How to train a neural coreference model— Neuralcoref 2](#)
- [6] [How to train and modify the neural coreference model](#)

4 Anexă A - Scenarii de utilizare

Legendă: *I = informație deținută de agent* *U = Cerere adresată de utilizator*
A = răspunsul agentului

- 1 I: Mrs. Robinson PROPER's phone NOMINAL is 0725555566, her PRONOMINAL office room is 236. She PRONOMINAL will attend the AI Conference in Bucharest on 20th February 2019. Today is 19th February 2019.
- U: What is the telephone number of Msr. Robinson ?
- A: The telephone number of Msr. Robinson PROPER is 0725555566.
- U: What about her PRONOMINAL office room?
- A: Her PRONOMINAL office room is 236.
- U: Will she PRONOMINAL attend the AI Conference NOMINAL tomorrow afternoon?
- A: Yes, she PRONOMINAL will.
- U: Where will it PRONOMINAL take place?
- A: It PRONOMINAL will take place in Bucharest.
- 2 I: My dog NOMINAL's name is Sudo. Him PRONOMINAL has to be fed in the morning and in the evening.

U: What's my dog **NOMINAL**'s name?

A: His **PRONOMINAL** name is Sudo.

U: When do I have to fed him **PRONOMINAL**?

A: In the morning and in the evening.

3 I: On Monday I am meeting Andreea **PROPER**. She **PRONOMINAL** will borrow me a book.

U: Who am I meeting on Monday?

A: You will be meeting Andreea **PROPER**.

U: What will She **PRONOMINAL** borrow me?

A: She **PRONOMINAL** will borrow you a book.

4 I: A good movie **NOMINAL** is Bohemian Rhapsody. It **PRONOMINAL** was directed by Bryan Singer. Rami Malek **PROPER** plays Freddie Mercury in the movie and he **PRONOMINAL** won an Oscar for Best Actor.

U: Recommend me a good movie.

A: A good movie **NOMINAL** is Bohemian Rhapsody.

U: Who directed it **PRONOMINAL**?

A: Bryan Singer directed it **PRONOMINAL**.

U: Who plays Freddie Mercury in the movie?

A: Rami Malek **PROPER** plays Freddie Mercury in the movie.

U: Did he **PRONOMINAL** win an Oscar?

A: Yes, he **PRONOMINAL** won an Oscar for best actor.

5 I: Tomorrow I have a Machine Learning course from 9 am to 11 am.

U: Set an alarm at 8 am.

A: Alarm set.

U: What course do I have tomorrow?

A: You have a Machine Learning **PROPER** course **NOMINAL**.

U: When will the course **NOMINAL** start?

A: It **PRONOMINAL** will start at 9 am.

U: How many hours does it **PRONOMINAL** last?

A: It **PRONOMINAL** lasts 2 hours.

6 I: Sherlock **PROPER** is a British crime drama television series based on Sir Arthur Conan Doyle's Sherlock Holmes detective stories. Created by Steven Moffat and Mark Gatiss, it **PRONOMINAL** stars Benedict Cumberbatch as Sherlock Holmes and Martin Freeman as Doctor John Watson. Thirteen episodes have been produced, with four three-part series airing from 2010 to 2017, and a special episode that aired on 1 January 2016.

I: Criminal Minds **PROPER** is an American police procedural crime drama television

series created and produced by Jeff Davis. It PRONOMINAL premiered on September 22, 2005, and has run for fourteen seasons on CBS. It PRONOMINAL tells the story of a group of behavioral profilers who work for the FBI PROPER as members of its PRONOMINAL Behavioral Analysis Unit (BAU).

U: Recommend me a crime television series.

A: Sherlock PROPER and Criminal Minds PROPER LIST are crime television series.

U: Which one of them PRONOMINAL is a British television series?

A: Sherlock PROPER is a British television series.

U: Who stars in it PRONOMINAL ?

A: Benedict Cumberbatch stars in it PRONOMINAL .

5 Anexă B - Rezultate obținute folosind NeuralCoref

1 Coreferințe identificate:

I: [Mrs. Robinson's phone: [Mrs. Robinson's phone, her, She]]

CONV: [Mrs. Robinson: [Mrs. Robinson, Her, her, Her, she, she], her office room: [her office room, Her office room]]

Coreferințe neidentificate:

What is the telephone number of Mrs. Robinson? Her phone number is 0725555566.

What about her office room? Her office room is 236. Will she attend the AI Conference NOMINAL

tomorrow afternoon? Yes, she will. Where will it PRONOMINAL take place? It PRONOMINAL

will take place In Bucharest.

2 Coreferințe identificate:

I: [My dog's name: [My dog's name, He]]

CONV: [my dog's name: [my dog's name, His name], His: [His, him]]

Coreferințe neidentificate: -

3 Coreferințe identificate:

I: [Andreea: [Andreea, She]]

CONV: [Andreea: [Andreea, she, She]]

Coreferințe neidentificate: -

4 Coreferințe identificate:

I: [A good movie: [A good movie, It, the movie], Rami Malek: [Rami Malek, he]]

CONV: [A good movie: [A good movie, the movie, the movie], it: [it, it], Freddie Mercury: [Freddie Mercury, Freddie Mercury], Rami Malek: [Rami Malek, he, he]]

Coreferințe neidentificate:

A good movie **NOMINAL** is Bohemian Rhapsody. Who directed it **PRONOMINAL** ?

Bryan Singer directed it **PRONOMINAL** .

5 Coreferințe identificate:

I: -

CONV: [a Machine Learning course: [a Machine Learning course, the course, It, it, It]]

Coreferințe neidentificate: -

6 Coreferințe identificate:

I: [Sherlock: [Sherlock, it]]

I: [Criminal Minds: [Criminal Minds, It, It], the FBI: [the FBI, its]]

CONV: [Sherlock and Criminal Minds: [Sherlock and Criminal Minds, them], Sherlock: [Sherlock, it, it]]

Coreferințe neidentificate: -