

Stanford CoreNLP Coreference Resolution

Coreference resolution: What is it? Some definitions	1
And why is it important?.....	1
Coreference Resolution with Stanford CoreNLP	2
Two types of coreference: nominal & pronominal	2
Four types of pronouns.....	2
Manual coreference.....	3
Blue and red: What do these colors mean?	3
Edit on the right-hand side.....	4
Reminders.....	4
References.....	5

Coreference resolution: What is it? Some definitions

Coreference resolution is the process of linking references to the same entity in a document. For example, in this sentence:

"I voted for Nader because he was most aligned with my values," she said.

(image © <https://nlp.stanford.edu/projects/coref.shtml>)

The red expressions (I, my, and she) refer to the same entity, and the blue expressions (Nader and he) refer to a different entity. It is necessary to determine which expressions refer to which entities for many different NLP tasks, such as summarization, question answering, and information extraction.

Coreference resolution is part of a general problem known in English grammar as **anaphora**, the use of a pronoun or other linguistic unit to refer back to another word or phrase. A word (a pronoun in the case of coreference pronominal resolution) that gets its meaning from a preceding word or phrase is called an **anaphor**. The preceding word or phrase is called the **antecedent**, **referent**, or **head**. Anaphora comes from the Greek word meaning "carrying up or back."

And why is it important?

Coreference resolution is an important first step for the accuracy of other NLP tasks. For instance, it is one of the first data cleaning steps involved in the SVO extraction pipeline, in order to have consistent subjects and objects. Without coreference resolution, a frequency distribution of subjects or objects, for instance, may give you a list of “he” “she” “they” that may refer to completely different entities.

Coreference Resolution with Stanford CoreNLP

Stanford CoreNLP offers three different approaches to coreference:

1. Deterministic (fast rule based)
2. Statistical (machine learning requiring dependency parsing)
3. Neural network (most accurate and slowest)

The NLP Suite relies on the neural network approach.

The script replaces all expressions referring to the same entity in a text with one representative expression.

For example, the following sentence:

“Bill Cato attempted to assault Mrs. Vickers, but her husband stopped him.”

Would become:

“Bill Cato attempted to assault Mrs. Vickers, but Mrs. Vickers’s husband stopped Bill Cato.”

Two types of coreference: nominal & pronominal

CoreNLP approaches implement both **pronominal** (i.e., pronouns referring to nouns, e.g., Barack Obama came to Boston; *he* said that...) and **nominal** (i.e., nouns referring to other nouns, e.g., Barack Obama came to Boston; *the President* said that...) coreference resolution. The algorithms do NOT resolve adverbial coreference (i.e., adverbs referring to nouns, e.g., Barack Obama came to Boston; *there* Obama said that...).

The NLP Suite implementation of coreference filters out the nominal coreference and focuses on the **pronominal coreference**. Too many errors otherwise.

Coreference resolution is still far from accurate, with perhaps 65% success.

Four types of pronouns

Pronominal resolution resolves four types of pronouns for the four different cases: nominative, possessive, objective, and reflexive.

The **nominative case** is used when the pronoun is the subject of the sentence. The nominative form pronouns are:

I, you, he/she, it, we, they.

The **possessive case** is used to show ownership or possession of something. The possessive form pronouns are:

My, mine, our(s), his/her(s), their, its, and yours.

The **objective case** is used as the direct object, indirect object, or the object of the preposition. The objective form pronouns are:

Me, you, him, her, it, and them.

Reflexive pronouns are words ending in -self or -selves that are used when the subject and the object of a sentence are the same (e.g., I believe in myself). They can act as either objects or indirect objects. The reflexive pronouns are:

myself, yourself, himself, herself, oneself, itself, ourselves, yourselves, and themselves.

Manual coreference

Due to the relatively low success rate of coreference resolution, the NLP Suite also implements a manual approach on the coreferenced output, displaying the original and coreferenced documents on two panels, side-by-side, **original on the left** and **coreferenced on the right** with the relevant pronouns and coreferences highlighted.

Comparing coref results (LEFT: ORIGINAL text; RIGHT: COREFED text). In BLUE pronouns NOT corefed; in RED pronouns corefed. EDIT text on the right and Save (or Quit). Use FIND bar to search text. — □ ×

Find	Original Text	Corefed Text
	There was an old sow with three little pigs, and as she had not enough to keep them, she sent them out to seek their fortune.	There was an old sow with three little pigs, and as she had not enough to keep them, she sent them out to seek their fortune.
	The first that went off met a man with a bundle of straw, and said to him: "Please, man, give me that straw to build me a house".	The first that went off met a man with a bundle of straw, and said to him: "Please, man, give me that straw to build me a house".
	Which the man did, and the little pig built a house with it.	Which the man did, and the little pig built a house with it.
	Presently came along a wolf, and knocked at the door, and said: "Little pig, little pig, let me come in".	Presently came along a wolf, and knocked at the door, and said: "Little pig, little pig, let me come in".
	To which the pig answered: "No, no, by the hair of my chiny chin chin".	To which the pig answered: "No, no, by the hair of my chiny chin chin".
	The wolf then answered to that: "Then I'll huff, and I'll puff, and I'll blow your house in".	The wolf then answered to that: "Then I'll huff, and I'll puff, and I'll blow your house in".
	So he huffed, and he puffed, and he blew his house in, and ate up the little pig.	So he huffed, and he puffed, and he blew his house in, and ate up the little pig.
	The second little pig met a man with a bundle of furze, and said: "Please, man, give me that furze to build a house".	The second little pig met a man with a bundle of furze, and said: "Please, man, give me that furze to build a house".
	Which the man did, and the pig built his house.	Which the man did, and the pig built his house.
	Then along came the wolf, and said: "Little pig, little pig, let me come in".	Then along came the wolf, and said: "Little pig, little pig, let me come in".
	"No, no, by the hair of my chiny chin chin".	"No, no, by the hair of my chiny chin chin".
	"Then I'll puff, and I'll puff, and I'll blow your house in".	"Then I'll puff, and I'll puff, and I'll blow your house in".
	So he huffed, and he puffed, and he puffed, and he huffed, and at last he blew the house down, and he ate up the little pig.	So he huffed, and he puffed, and he puffed, and he huffed, and at last he blew the house down, and he ate up the little pig.
	The third little pig met a man with a load of bricks, and said: "Please, man, give me those bricks to build a house with".	The third little pig met a man with a load of bricks, and said: "Please, man, give me those bricks to build a house with".
	So the man gave him the bricks, and he built his house with them.	So the man gave him the bricks, and he built his house with them.
	So the wolf came, as he did to the other little pigs, and said: "Little pig, little pig, let me come in".	So the wolf came, as he did to the other little pigs, and said: "Little pig, little pig, let me come in".
	"No, no, by the hair of my chiny chin chin".	"No, no, by the hair of my chiny chin chin".
	"Then I'll huff, and I'll puff, and I'll blow your house in".	"Then I'll huff, and I'll puff, and I'll blow your house in".
	Well, he huffed, and he puffed, and he huffed and he puffed, and he puffed and he huffed; but he could not get the house down.	Well, he huffed, and he puffed, and he huffed and he puffed, and he puffed and he huffed; but he could not get the house down.
	When he found that he could not, with all his huffing and puffing, blow the house down, he said: "Little pig, I know where there is a nice field of turnips".	When he found that he could not, with all his huffing and puffing, blow the house down, he said: "Little pig, I know where there is a nice field of turnips".
	"Where?"	"Where?"
	said the little pig.	said the little pig.
	"Oh, in Mr. Smith's Home-field, and if you will be ready to-morrow mor-	"Oh, in Mr. Smith's Home-field, and if you will be ready to-morrow mor-

Blue and red: What do these colors mean?

On the **left-hand side**,

pronouns cross-referenced by CoreNLP are tagged in **RED**.

pronouns **NOT** cross-referenced by CoreNLP are tagged in **BLUE**.

On the **right-hand side**,

pronouns cross-referenced by CoreNLP are tagged in **RED**, with the pronouns replaced by the referenced nouns.

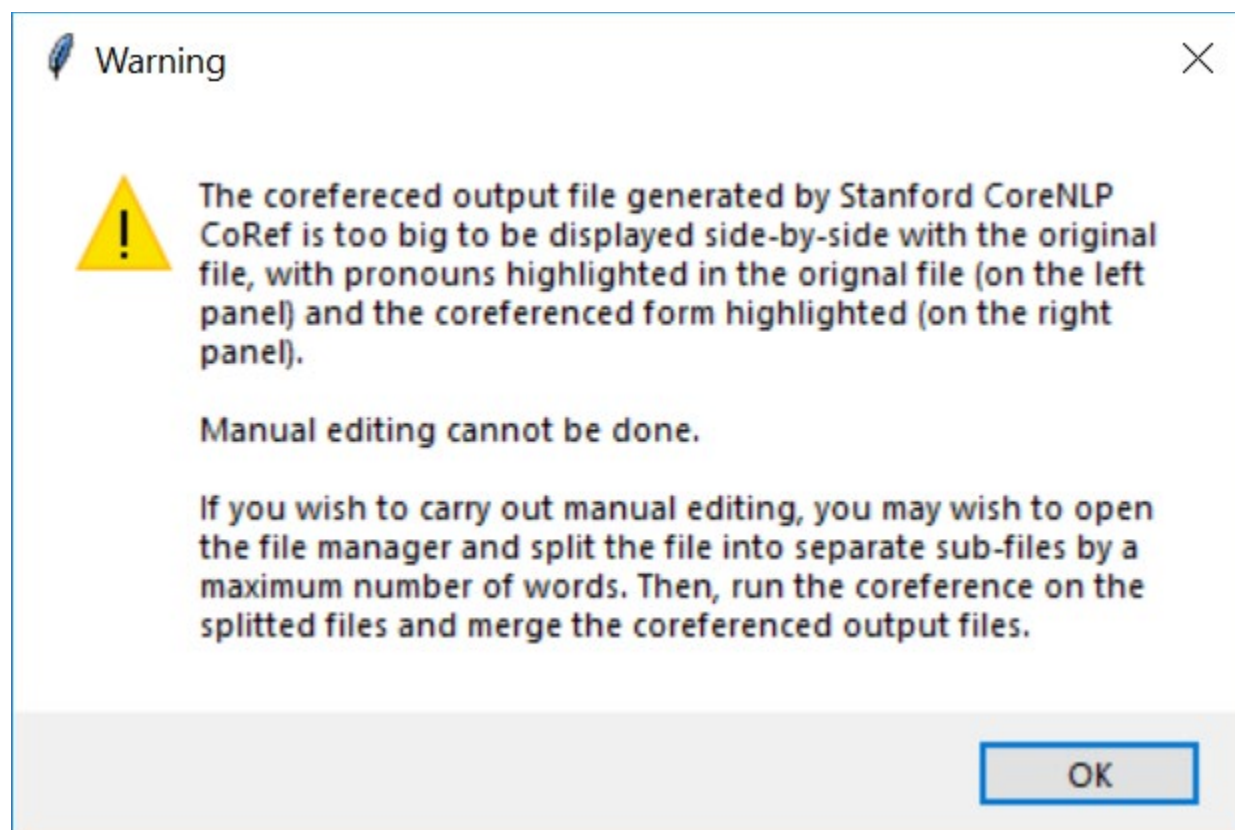
Although the highlighting is not perfect, it does provide the user an immediate visual tool of comparison.

Edit on the right-hand side

The user can edit any unresolved or wrongly resolved pronominal cases directly on the right panel, as if it were any text editor and then save the changes.

Reminders

Since the function works in memory, for large files memory this may not be an option. If that is the case, the script will warn the user.



The script similarly warns the user to deselect manual coreference when processing a directory.

References

- Lee, Heeyoung, Angel Chang, Yves Peirsman, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. 2013. “Deterministic coreference resolution based on entity-centric, precision-ranked rules.” *Computational Linguistics*, 39(4).
- Clark, Kevin and Christopher D. Manning. 2015. “Entity-Centric Coreference Resolution with Model Stacking.” In *Proceedings of the ACL*.
- Clark, Kevin and Christopher D. Manning. 2016. “Deep Reinforcement Learning for Mention-Ranking Coreference Models.” In *Proceedings of EMNLP*.
- Clark, Kevin and Christopher D. Manning. 2016. “Improving Coreference Resolution by Learning Entity-Level Distributed Representations.” In *Proceedings of the ACL*.