

Developing a sample
resource +
Inception: survival guide

*Linguistic Resources for Natural Language Processing
LM Language Technologies and Digital Humanities
2024-25*

Cristina Bosco

Development of a linguistic resource

The development of a linguistic resource involves several main **steps**:

- data collection
- definition of a format for data annotation
- application of the annotation on the data
- check and correction of the annotated data
- release of the resource

Development of a linguistic resource

The development of a linguistic resource involves several main **steps**:

**Assumed as solved:
we use a small set of sample
and UD format**

- data collection
- definition of a format for data annotation
- application of the annotation on the data
- check and correction of the annotated data
- release of the resource

Development of a linguistic resource

The development of a linguistic resource involves several main **steps:**

To be solved:

we apply UD using UDPipe and

- data collection **we correct the output using Inception**
- definition of a format for data annotation
- application of the annotation on the data
- check and correction of the annotated data
- release of the resource

Development of a linguistic resource

See Exercise on Development of linguistic resources (lesson 10).
It includes the sentences of the sample corpus.

Apply UDPipe on these sentences, save the output in a file.

Open the file with a visualisation tool to see the dependency trees.
Are they correct?

Correct the trees using Inception.

Create a project

Select Create

Give a name to the project

Import a file **selecting the format CONLL-U.**

Warning: If the format is not selected the file will be open as a plain text, no trees will be shown.

Delete a project (irreversible action!!!)

Open the project, select Setting in the left panel and Danger Zone at end of Settings. Write the name of the project and click Delete.

USER GUIDELINES:

<https://inception-project.github.io/releases/29.9/docs/user-guide.html>

Modify a PoS tag in Inception:



To modify the **Pos or features** click on it.
When selected it turns orange.

You can modify all in the right panel



Layer Morphological fe ▾

Span 🗑 Delete ✕ Clear

← →

Layer

Morphological features

Text ⚙

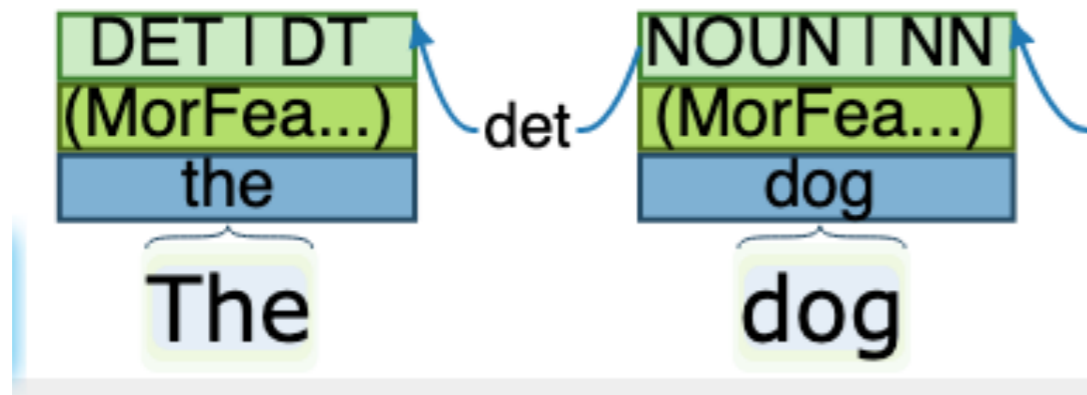
The

No links or relations connect to this annotation.

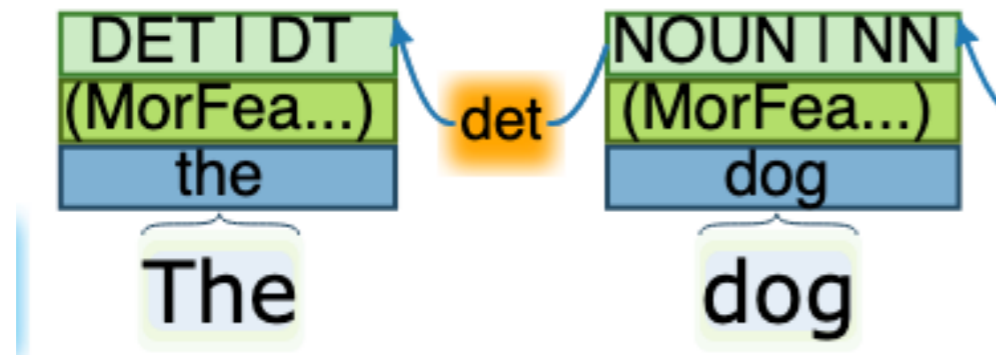
Features

Definite=Def|PronType=Art

Modify a relation in Inception:



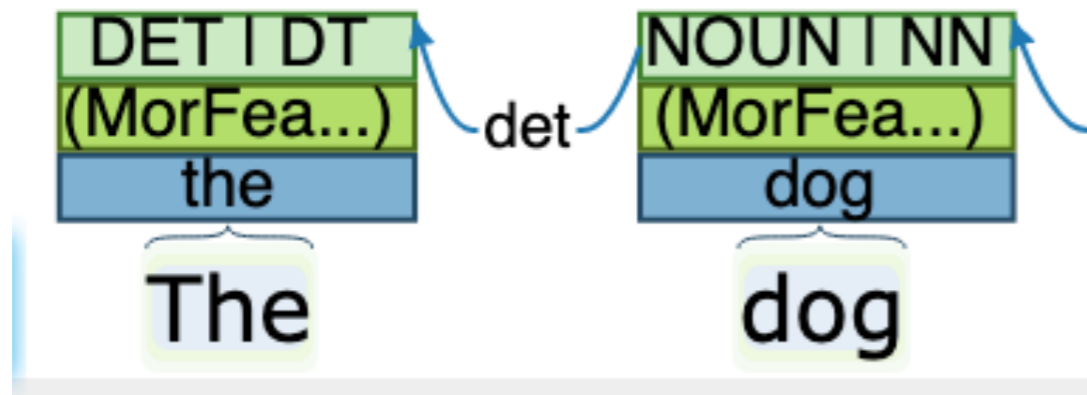
To modify the **label** (det) click on it.
When selected the label turns orange



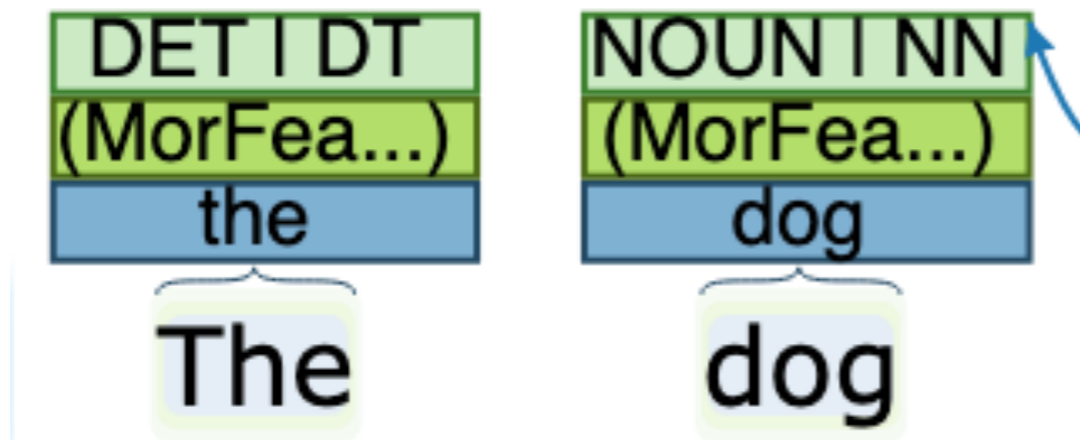
In the right panel find the relation and select the new name for it

The screenshot shows the Inception interface for editing a relation. At the top, there are buttons for "Delete", "Reverse", and "Clear". Below these are navigation arrows. The "Layer" is set to "Dependency". The "From" field contains "dog" and the "To" field contains "The". The "Flavor" section shows "basic" and "enhanced" options. The "Relation" dropdown menu is open, showing "det" selected. A red arrow points from the text below to the "det" option in the dropdown.

Modify a relation in Inception:

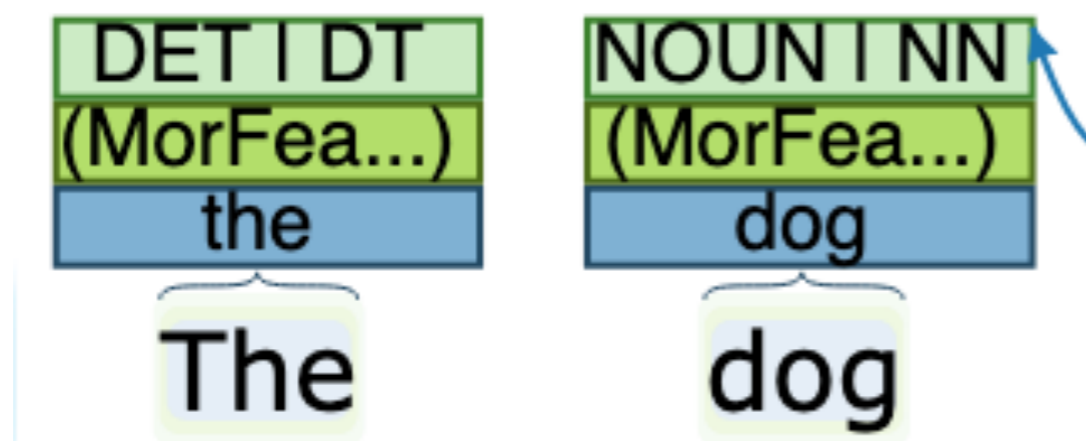


To **delete the relation** (det) click on it. Click on DELETE in the right panel



The screenshot shows the 'Relation' panel in the Inception interface. At the top, there are three buttons: 'Delete' (red), 'Reverse' (grey), and 'Clear' (grey). Below these are two arrows pointing left and right. The 'Layer' section contains a text input field with the value 'Dependency'. The 'From' section contains a text input field with the value 'dog'. The 'To' section contains a text input field with the value 'The'. The 'Flavor' section contains a list with two items: 'basic' and 'enhanced', each with an information icon. The 'Relation' section contains a dropdown menu with the value 'det' and a close button.

Modify a relation in Inception:



Layer **Part of speech** ▾

Create a dependency relation by drawing an arc between annotations of this layer.

Relation

Delete Reverse Clear



Layer

Dependency

To create a relation between a **head and a dependent** select in the right panel the layer Part of Speech. Click on the head and drag up to the dependent.

A relation with the generic label will be created that you can change.

