

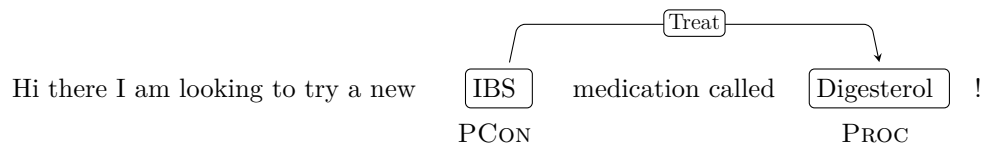
Annotation Guidelines for Condition-Treatment Relation Extraction

Sichang Tu, Jinho D. Choi

April, 2022

1 Project Overview

This project is a collaborative project between [Emory NLP](#) and [Real Life Science](#), aiming to build an NLP pipeline that recognizes all entities (e.g., ‘PATIENT CONDITION’ and ‘PROCEDURE’) in disease specific social media data. For instance, in the following example, ‘IBS‘ is the condition that the patient/poster has, and ‘Digesterol‘ is the medicine that the patient is currently taking. Our goal is to capture all disease-related information in the data to help the patient, the caregiver, and the health care practitioner.



2 Annotation Scheme

The current guideline has two layers: **Entity** and **Relation**. For **Entity** layer, we have 3 categories of labels: **CONDITION**, **TREATMENT**, and **ATTRIBUTE**. **CONDITION** can be further divided into three entity labels, namely **PATIENT CONDITION**, **CAREGIVER CONDITION**, and **UNSPECIFIED CONDITION**, depending on who is the subject of the medical condition. **TREATMENT** contains two labels, **MEDICINE** and **PROCEDURE**.

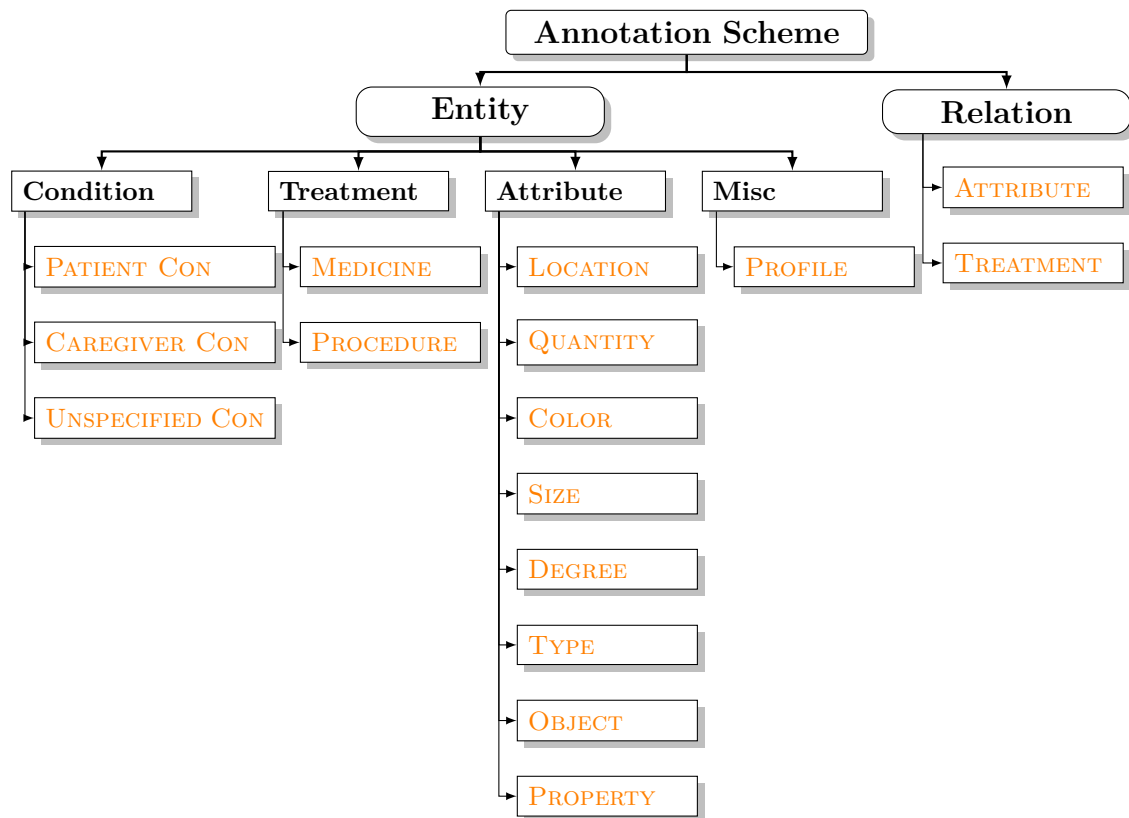


Figure 1: The structure of the annotation scheme.

3 Entity

3.1 CONDITION

Generally, **CONDITION** captures the disease and any related condition that the subject is dealing with. It can be divided into three subcategories as follows:

PATIENT CONDITION This label refers to the condition that the writer of the passage has or is dealing with. It could be a specific disease or any related symptoms, side effects, or impairment caused by a disease or medication.

Hi 2 years ago I was diagnosed with lupus .
PCon

CAREGIVER CONDITION Different from **PATIENT CONDITION**, **CAREGIVER CONDITION** is the condition affecting someone the writer of passage cares for (e.g., family members or friends).

I am the mother of a son who was diagnosed with tourette's at age 6.
CCon

UNSPECIFIED CONDITION It appears in the context where the subject of the condition is unclear or general. Usually the person that suffers from the condition does not fit into the **PATIENT** / **CAREGIVER CONDITION** labels (e.g., other users in the previous post threads).

im very sorry to hear about your diagnosis of PND .
UCon

Another case of **UNSPECIFIED CONDITION** happens when the condition is assumed or deduced. Since it is not the real condition that the patient actually has, we label it as **UNSPECIFIED CONDITION**. For example, in the following sentence, the patient is talking about a medical fact that gluten may cause damage to the small intestine, instead of some conditions the patient is experiencing.

I know that bread (gluten) can induce damage to the inner lining of the small intestine .
ATTR

3.2 TREATMENT

Generally, TREATMENT annotates any medical treatment (e.g., medicine, surgery, or even counseling) that the subject is taking. TREATMENT can be further divided into **MEDICINE** and **PROCEDURE**.

MEDICINE It refers to any substance used in treating disease or illness. It could be a drug name, a brand name or a type of medication.

There is an over the counter medication called Mucus Relief DM .
MED

PROCEDURE It refers to any medical procedure intended to achieve a result in the delivery of healthcare. Common kinds of procedures include therapeutic (i.e., intended to treat, cure, or restore function or structure), propaedeutic, surgical and anesthesia.


Diagnosed with the disease in 2002, tried lumpectomy and chemo .
PROC PROC

Note that TREATMENT can exist without CONDITION in the same post.

I was given a prescription called “ magic mouthwash ” by my Oncologist.
MED

However, if there exists CONDITION labels in the text, we always attach the TREATMENT to the nearest corresponding CONDITION using ingoing relation.

I was dx'd with anxiety panic disorder and put on zoloft .
PCON MED



TREATMENT also includes any non-traditional or ill-advised treatments like alcohol or hard drugs.

Dude, you medicate with Whisky ?
MED

Any diagnostic procedures (e.g., ‘biopsy’ or ‘ct scan’) are not considered as treatment, and thus the statistical results should not be annotated as a condition. For instance, phrases like ‘muscle biopsy’ should not be labeled since they are diagnostic tests for the disease instead of treatments.

- (1) My Neuromuscular specialist gave me a muscle biopsy and made a final diagnosis of the disease.

The scope of ‘Treatment’ also excludes diet and fitness. The ‘nutrisystem’ is very misleading since it seems a medical term. However, it is a website, Nutrisystem (misspelled as ‘nutrisysten’), which delivers healthy prepared meals. Please search the word online if you are not sure what it means.

- (2) I have lost weight last year because I began to work out.
- (3) I have changed my diet again to nutrisysten.

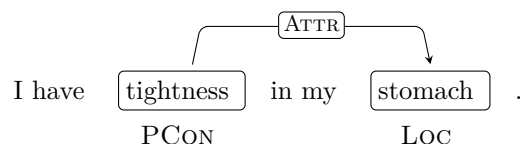
Besides, the time span and dose for a treatment should also be excluded.

They put me in the psy unit for 5 days.
PROC

The doctor gave me .15mg levothroid .
MED

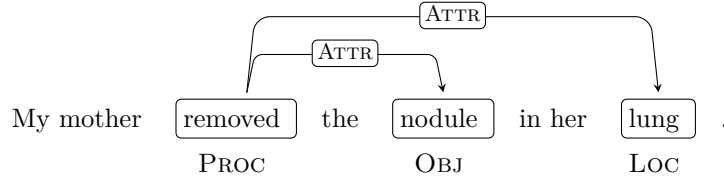
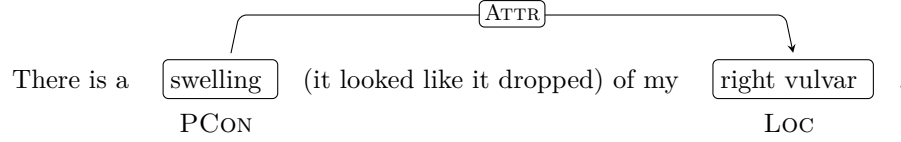
3.3 ATTRIBUTE

A CONDITION or TREATMENT may have modifiers, usually adjectives or nouns used attributively, to describe them. Each modifier should be attached to the CONDITION / TREATMENT by an ingoing relation. For example, in the following sentence, ‘tightness’ should be labeled as **PATIENT CONDITION**, with a link going outward to its **LOCATION** modifier. Note that ATTRIBUTE labels can not exist without CONDITION / TREATMENT labels within the same sentence.

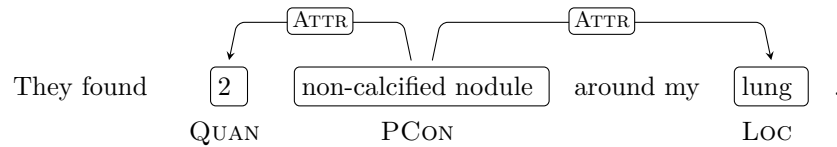


Currently, we have the following attributes:

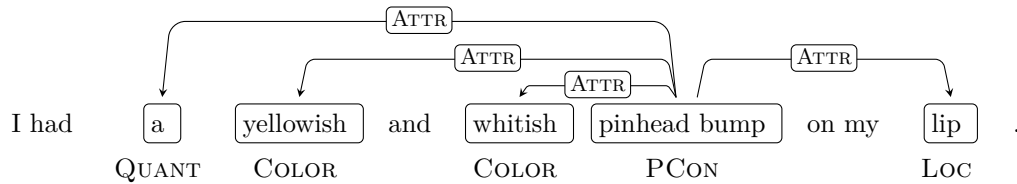
LOCATION It describes where the **CONDITION** is located or where the **TREATMENT** happens, such as body parts and organs. It can be singletons (single word) or noun phrases.



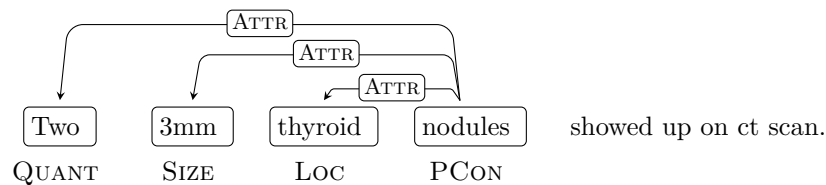
QUANTITY Sometimes quantity modifier is used when patient describing their conditions, especially symptoms. In order to capture this information, we label the quantity determiner as **QUANTITY**.



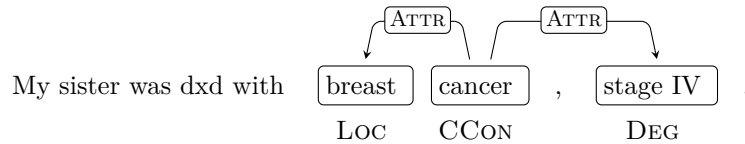
COLOR As the label indicates, **COLOR** refers to the modifiers that describes the color of the condition.



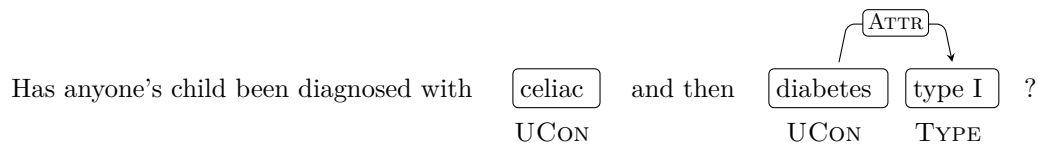
SIZE **SIZE** is also important information to capture when labeling patient's condition. It describes the magnitude and dimension of a condition.



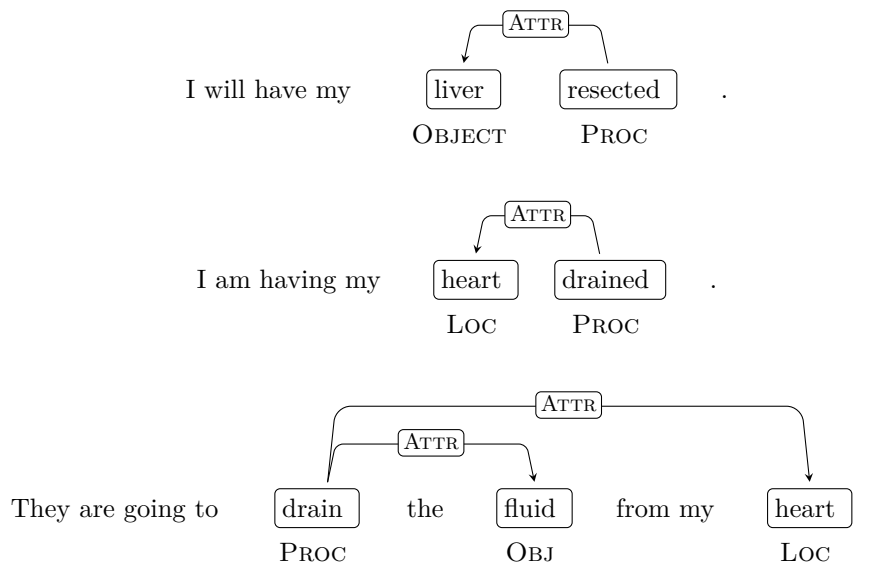
DEGREE **DEGREE** shows how severe the condition is, such as disease stages that provides information on disease development.



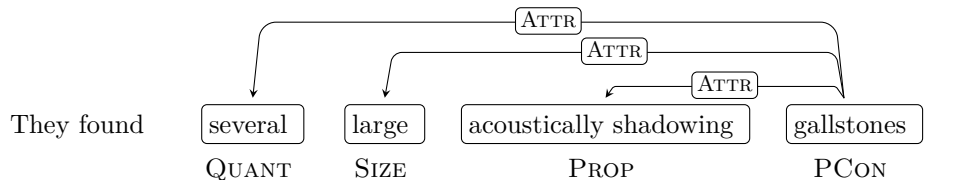
TYPE Some diseases may have various types with different symptoms. For example, diabetes has three different types, including Type 1, Type 2, and Gestational.



OBJECT This label annotates any object that the **TREATMENT** is directed. Sometimes **OBJECT** is difficult to distinguish from **LOCATION**. Taking the following three sentences as examples, the ‘liver’ in the first sentence is labeled as **OBJECT** since the thing to be resected is a part of the ‘liver’. However, in the second sentence, ‘heart’ is a **LOCATION** since the exact thing to be drained is actually fluid in the heart, rather than the organ itself. Hence, the **OBJECT** in sentence 2 is implicit. And it is equal to sentence 3, which is the explicit version of sentence 2.



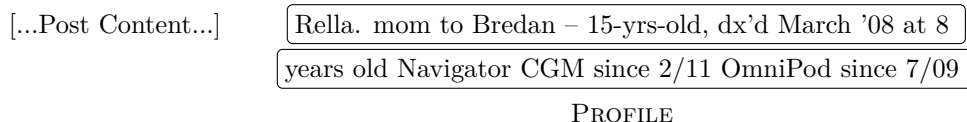
PROPERTY CONDITION / TREATMENT sometimes may have modifiers that do not fit into the above categories. In this case, we use the **PROPERTY** label to capture the modifiers that describe other properties or characteristics of the entity.



3.4 Miscellaneous

MISCELLANEOUS covers any other entity that does not fit in any of the previous categories, and that may be useful to the condition-treatment extraction.

PROFILE Sometimes, additional information is provided after the post. As shown in the example below, the additional information usually follows the post content. Rather than a complete sentence, it always starts with the username of the poster and lists information including the condition of the patient, devices used by the patient, and the treatment taken by the patient. Since it is not a grammatical sentence, it cannot fit in the CONDITION/TREATMENT labels. Hence, we adopt **PROFILE** to label the additional information as a whole.

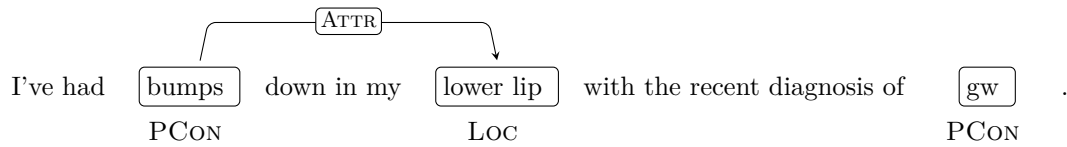


4 Relations

The Relation layer is used to annotate the relation between two entity spans. We currently have two edge labels: **ATTRIBUTE** and **TREATMENT**.

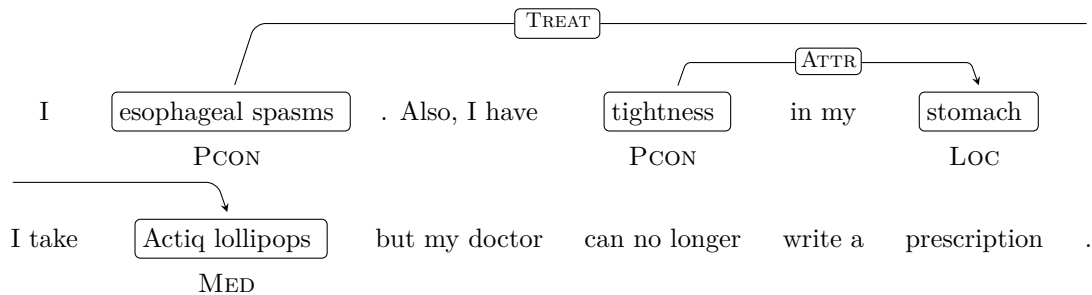
4.1 ATTRIBUTE

ATTRIBUTE captures the relation between CONDITION / TREATMENT labels and their ATTRIBUTE labels. The direction of the edge is always from CONDITION / TREATMENT to ATTRIBUTE. Since ATTRIBUTE labels must be in the same sentence with CONDITION / TREATMENT labels, the relation is not a cross-sentence relation.



4.2 TREATMENT

TREATMENT captures the relation between **CONDITION** labels and **TREATMENT** labels. The direction of the arc is always from **CONDITION** to **TREATMENT**. Note that we should attach the **TREATMENT** to the nearest corresponding **CONDITION**. The **TREATMENT** relation might be cross-sentence relation, since all not **TREATMENT** labels are in the same sentence with its corresponding **CONDITION** labels. In the following example, the patient has two conditions. ‘esophageal spasms’ and ‘tightness’. The medicine ‘Actiq lollipops’ is actually the treatment for ‘esophageal spasms’. Hence, we should connect the medicine to the corresponding condition with an ingoing arc.



5 General Annotation Conventions

5.1 General Terms

When labeling **CONDITION**, general terms such as ‘symptoms’, ‘problems’, or ‘disease’ should be excluded unless:

1. They are parts of the disease names. For example, in the first sentence, ‘symptoms’ should be excluded. However, in the second sentence, ‘common’ should be labeled since it is a part of the disease name.

I don't know if it's the MS symptoms.

PCON

I was diagnosed with cvid common variable immunodeficiency .
PCon PCon

2. They have modifiers or descriptors before them. That is to say, if a general term shows up in the text as a singleton without any modifier or descriptor, do not label it. As shown in the following example, ‘autoimmune disease’ is labeled since the ‘disease’ has ‘autoimmune’ as the modifier to specify what kind of disease those people have. But the singleton ‘disease’ at the end of the sentence is not labeled.

Lots of ppl here are dxd with autoimmune disease .
UCon

As for TREATMENT, we should not annotate a general reference in the post without any modifiers or descriptor indicating the specific medication or treatment, such as ‘meds’ and ‘pills’. For instance, the ‘pills’ in sentence (1) is not labeled, while ‘birth control pills’ in sentence (2) is labeled since it specifies the pill is used for birth control.

I think I can take some pills.

In this case, people can take birth control pills .
MED

5.2 Annotation Consistency

Consistency is an important element that influences annotation quality, showing how we follow the guideline. Similar cases and phenomena across the dataset should receive consistent analysis. Hence, when we annotate one entity in the post, we should also label all other references. For example, in the following sentence, the occurrences of ‘pus’ should be both labeled as **PATIENT CONDITION**. And it would be considered inconsistent if we only label the first ‘pus’.

I had pus drainage incision with a scalpel so the who pus
PROC PCon

can get out. Four days ago the doctor said that
there isn't any pus already.
PCon