Annotation Guide - TRANSMATT

13/05/2020 LENTSCHAT Martin

This document is intended to guide the annotator in the use of the WebAnno tool. The version used here is made available by the ICO team of UMR IATE and can be found at: https://ico.iate.inra.fr/webanno/login.html.

As an annotator you have to log in with the following login and password:

login firstname.name

password firstname.name

Using the Anno web tool

Once logged in you can access the annotation function from the main menu. By clicking on the button **Annotation** a new window opens and you can then select one of the projects you are involved in (left column) and then the document you wish to work on (right column).

The colour of the document names indicates the status of the work on the document:

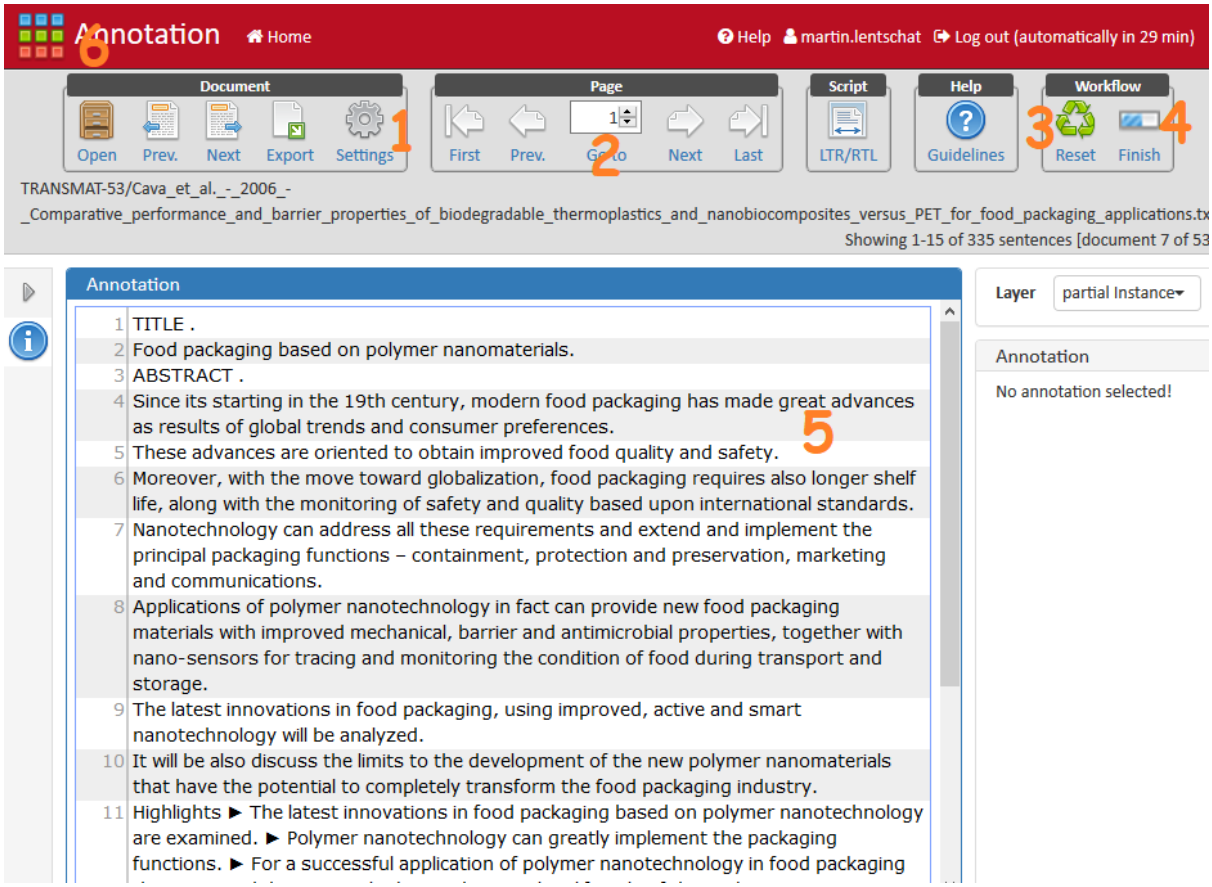
black annotation not started

blue annotation in progress

red annotation completed

The annotation interface and its functionalities are:

1. **Settings** allows you to adjust the display of the **text**
2. buttons in the **page** tab allow you to quickly navigate through the text.
3. the **reset** button deletes all annotations on the active document.
4. click the **finish** button when you consider that you have completed the annotation.
5. this is where you will make the annotation of the text.
6. clicking on the **logo** allows you to return to the home page.

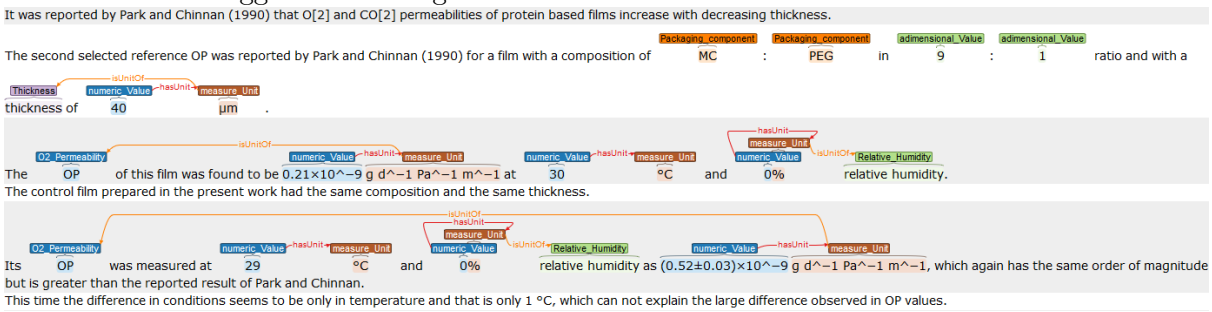


Here we seek to annotate scientific articles by locating certain *Aguments* involved in *N-Ary Relations* of permeability. This therefore concerns several types of information present throughout the texts.

Annotation creation

To create an annotation, simply double-click a word or highlight a sequence of word. This selection is made down to the character to be as precise as possible. The *Label* is to be selected in the right tab once the annotation has been created.

Some *Labels* will need to be linked together by connections. These are created by "drag and drop" from the selected annotation to another. Pay attention to the direction of the connection. These linked *Labels* are not necessarily in the same sentence. This is not a problem because a connection can be dragged over a long distance.

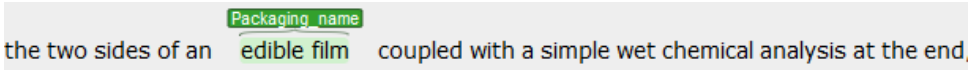


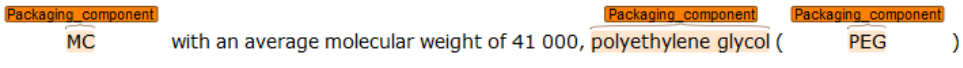
Once an annotation has been created, it can be modified or deleted by double-clicking on it again.

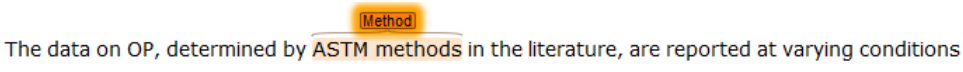
The following sections detail the selection criteria for *Labels* and connections.

Labels and Symbolic Argument connections

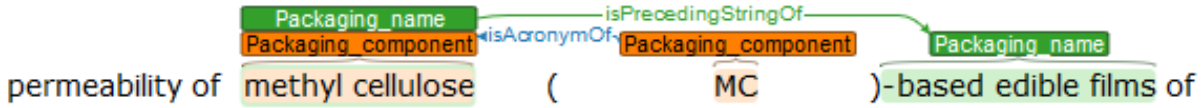
The instances of symbolic arguments to be annotated will consist of the terms (i.e. words or sequences of words) relating to food packaging and the method used to measure permeability. We will distinguish the name used by the authors to designate and categorise a package from the components used to create it.

`Packaging_Name` 

`Packaging_Component` 

`Method` 

As these terms are " self-supporting ", they are rarely in need of connections (with the exception of component proportions, see below). The connection `isAcronymOf` may be used (this job is optional), to indicate links between terms and their acronyms. The connection `isPrecedingStringOf` allows you to link annotations of the same *Labels*, to indicate that they are the same term. It is also possible to overlay annotations if justified.

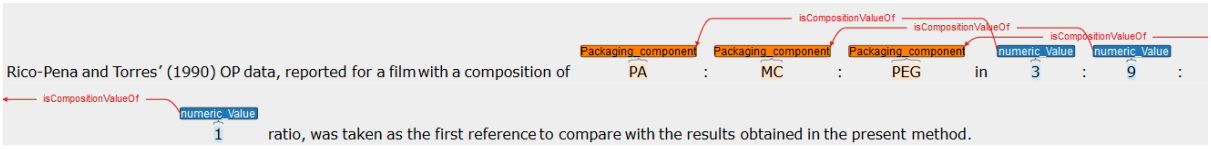


Packaging Components and connections with their proportions

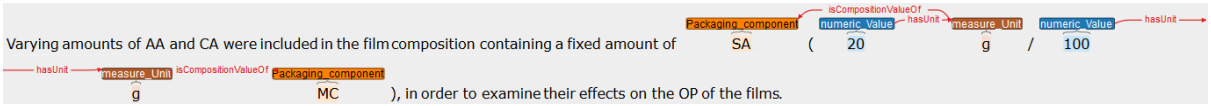
There are two scenarios in the documents for the quantities of packaging components: ratios and mass quantities. Two elements are always present (a numerical value and a packaging component), sometimes with a unit of measurement. The connection used to link component and numerical value or unit of measurement, when available, is `isCompositionValueOf`.

Example :

- `Packaging_Component` and `numerical_Value`



- `Packaging_Component`, `numerical_Value` and `measure_Unit`



Labels and Quantity Argument connections

The instances of quantity arguments to be annotated will be more complex, made up of multiple *Labels*. They always consist of a numerical value (`numeric_Value`) with a unit of measurement (`measure_Unit`) and sometimes a term indicating its dimension :

`Thickness` thickness of a food packaging, e.g. : *"thickness", "thick" ...*

`Temperature` control temperature used during permeability readings, ex : *"temperature", "T", "temp." ...*

`Relative_Humidity` control humidity of permeability measurements, e.g. : *"relative humidity", "RH" ...*

`Relative_Humidity_Difference` similar to the previous one, present when measuring water vapour permeability (H2O_Permeability)

`Partial_Pressure_Difference` difference in pressure, ex : *partial pressure difference*

`CO2-H2O-O2_Permeability` term of permeability, e.g. : *"oxygen permeability", "WVP", "[CO2] perm" ...*

Warning : Values not involved in permeability relation such as the temperature for preparing the packaging, the storage humidity etc. are not concerned.

Connections indicate that different elements constitute the same instance of arguments. There are several of them:

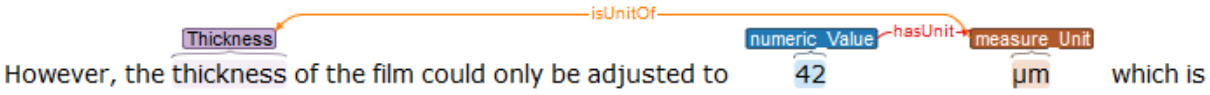
`hasUnit` goes from `numeric_Value` and link it to `measure_Unit`

`isUnitOf` goes from `measure_Unit` and link it to one of the quantity term

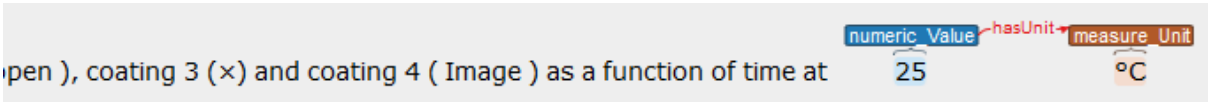
The presence of a term giving the dimension of the numerical value is important in order to disambiguate it as much as possible. It can sometimes be present not in the sentence but in the window (± 1 sentence) around the unit of measurement and the numerical value. Nevertheless, among the three elements ideally required, some are sometimes missing.

The possible cases for annotating the terms and their connections are therefore :

1 The numerical value is linked to the unit of measurement, which in turn is linked to the quantity term.



2 The numerical value is linked to the unit of measurement, the quantity term is missing.



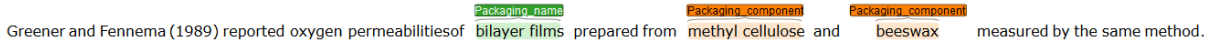
The case of Citations

Authors of an article sometimes refer to work outside the article in order to compare their work. In this case it may appear that arguments are present, often relating to the packaging being discussed.

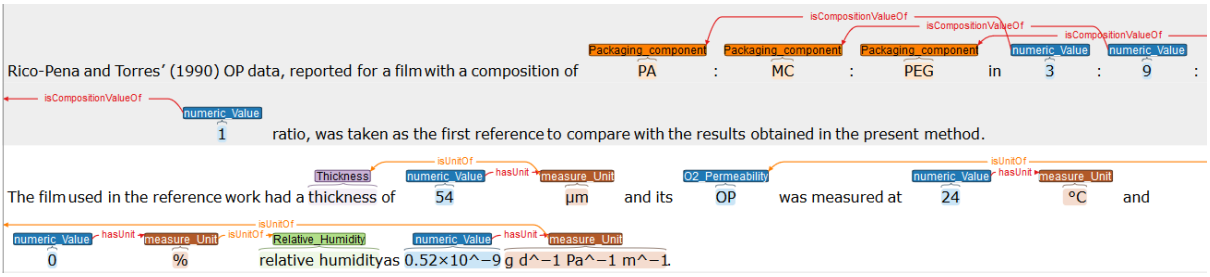
Warning : not being the focus of the paper, the information presented is often incomplete. Annotate only if several arguments are present, `Packaging` and `Permeability` being mandatory.

Example :

- Invalid



- Valid



The case of Tables

Tables are present in the documents. These are composed of a legend, note or quotation and a table. The latter is easily recognizable because each line of the table is then preceded by the word *"Line"* and *pipes* (*"—"*) are present in order to differentiate the cells.

Given the complexity for the annotator, as well as for the annotation client, to annotate tables I recommend **to annotate only text lines** as the caption and **do not annotate the contents of table cells**.

Recommandations

- take the time to understand what you are reading and do several readings. The text surrounding an item often tells you whether an annotation is warranted.
- if you do not know whether a word should be included in an annotation or not then annotate the broadest one.
- annotate according to what you think is right, not what you think is consensual.
- if a case makes you doubt and is not covered by the guide, then do not annotate.
- the same information that appears many times in the text (such as the name of a package) may be annotated only once or twice. There is no need to repeat an annotation that you are certain you have already created.