## **CAT User Manual for the NewsReader EU Project**

Technical Report NWR-2014-5 Version FINAL

Giovanni Moretti, Rachele Sprugnoli

Fondazione Bruno Kessler moretti, sprugnoli@fbk.eu





### BUILDING STRUCTURED EVENT INDEXES OF LARGE VOLUMES OF FINANCIAL AND ECONOMIC DATA FOR DECISION MAKING

ICT 316404

## Contents

1.	Introduction to CAT: an Overview	5
2.	Running CAT	5
3.	Using CAT	
3	.1 Exploring the Interface	
3	.2 Creating a Corpus	
	.3 Adding Files to a Corpus	
	.4 Importing the NewsReader Annotation Task	
	.5 Modifying an Existing Markable	
	.6 Modifying an Existing Relation	
	3.6.1 Modifying Markable and Relation Attributes	
3	7 Opening a File	
	8 Performing Markable Annotation	
	3.8.1 Starting the annotation	15
	3.8.2 Annotating a markable	
	3.8.3 Annotating attributes of a markable	
	3.8.4 Modifying/deleting an annotation	
3	9 Annotating a Relation	
	3.9.1 Changing or Deleting the Source or the Target of a Relation	
	3.9.2 Deleting a Relation	
	3.9.3 Annotating Relation Attributes	
3	.10 Annotating Empty Tags	
	.11 Hiding Annotated Markables	
	.12 Saving a File	
	.13 Exporting a Corpus	
4.	Calculating Inter-Coder Agreement	23
5.	List of Keyboard Shortcuts in Alphabetical Order	23
6.	Updates from the previous version of the manual	24
Ref	erences	

## List of Tables

FIGURE 1: THE APPLICATION MENU BAR	6
FIGURE 2: WINDOW WITH THE NUMBER OF ANNOTATED MARKABLES IN A FILE	7
FIGURE 3: THE TEXT PANEL MENU BAR	8
FIGURE 4: THE RELATION PANEL	
FIGURE 5: THE LAYER PANEL	9
FIGURA 6: CAT INTERFACE	9
FIGURE 7: HIDE/UNHIDE THE CORPUS PANEL FIGURE 8: HIDE/UNHIDE THE RELATION PANEL	10
FIGURE 8: HIDE/UNHIDE THE RELATION PANEL	10
FIGURE 9: WORKING ENVIRONMENT WITH A REDUCED WIDTH OF THE TEXT PANEL	11
FIGURE 10: IMPORTING A TASK	
FIGURE 11: THE TASK SELECTION DROP-DOWN BOX	
FIGURE 12: THE DEFINE MARKABLE WINDOW	
FIGURE 13: CHANGE THE ORDER OF THE ATTRIBUTE VALUES	
FIGURE 14: ICONS OF THE DIFFERENT TYPES OF ATTRIBUTES	
FIGURE 15: EXAMPLE OF AN ATTRIBUTE WINDOW WITH UNSAVED VALUES (IN RED)	
FIGURE 16: EXAMPLE OF AN ATTRIBUTE WINDOW WITH SAVED VALUES (IN BLACK)	
FIGURE 17: ASSIGNING THE VALUE TO A REFERENCE LINK ATTRIBUTE	
FIGURE 18: CHOOSE THE MARKABLE TO MODIFY	
FIGURE 19: CHOOSE THE TYPE OF RELATION TO ANNOTATE	
FIGURE 20: CREATE A NEW RELATION	19
FIGURE 21: ADD SOURCE AND TARGET OF A RELATION	
FIGURE 22: DRAG AND DROP THE SOURCE TO THE TARGET OF THE RELATION	
FIGURE 23: SOURCE AND TARGET HIGHLIGHTED IN THE TEXT	
FIGURE 24: DELETE ONE SOURCE OR ONE TARGET IN A RELATION	
FIGURE 25: EMPTY TAGS	22

## 1. Introduction to CAT: an Overview

The Content Annotation Tool, hereafter mentioned with the acronym CAT, is a general-purpose web-based text annotation tool created by the Center for the Evaluation of Language and Communication Technologies (CELCT) and now maintained and developed by Fondazione Bruno Kessler (FBK). The aim of the tool is to make the annotation activity as intuitive as possible, supplying, at the same time, a rich set of features. The main strengths of CAT are flexibility, practical usability and customizability. It also supports multi-layer annotation in order to combine the annotation of several linguistic/semantic layers, it provides a standardized XML stand-off output format to help convertibility.

• Lenzi, V. B., Moretti, G., & Sprugnoli, R. (2012). CAT: the CELCT Annotation Tool. In LREC (pp. 333-338).

## 2. Running CAT

No installation is required given that CAT runs in your browser: you only need a supported browser (see Table 1) and the Flash plugin properly installed (<u>http://get.adobe.com/it/flashplayer/</u>). To obtain access to CAT go to the following URL: <u>http://dh.fbk.eu/resources/cat-content-annotation-tool</u>.

A local application has been development as well. To obtain it please send an email to <u>CAT.support@fbk.eu</u>.

Browser	Visualization	Editing
Chrome	Full	Full
Safari	Full	Full
Firefox	Full	Almost Full (some shortcuts are
		not properly working)
Explorer		Not supported

The following table summarizes the current browser support for CAT functions.

### Table 1: supported browser

Please note that browsers often change their versions so some issues can arise: the above table is based on testing with recent browser versions, older or newer versions may not support the same features.

## 3. Using CAT

### **3.1 Exploring the Interface**

The main window of CAT appears on the screen after the login. The window is divided into three parts:

- 1) the Corpus Panel on the top left: corpora and files are listed here; the list can be hidden clicking on the eye icon
- the Layer Panel on the bottom left: annotation layers (i.e. the list of available markables) and empty tags (if present) are listed in two different tabs;
- 3) the large Text Panel on the right: text to be annotated are shown here.

On the top of the main window, the application menu bar is displayed (fig.1): it contains File, Edit, Task, Markable and Relation pull-down menus which can be selected passing the mouse over the bar.

File Edit Task Markable Relation Statistics Help	
--	--

Figure 1: the application menu bar

The **File** menu permits the following actions:

- New Corpus: creates a new corpus to be annotated; files of the corpus are displayed in the Corpus panel;
- Import: imports raw texts or tokenized files in a corpus;
- Import XML CAT file: imports one or more files already annotated and formatted following the CAT XML format in a selected corpus
- Export as standoff XML: exports the annotated corpus in the CAT XML standoff format;
- Logout: allows to exit from your CAT account

The **Edit** menu permits the following actions:

- Preferences: opens up a Preferences window in which you can customize some elements of CAT.

The **Task** menu permits the following actions:

- Define Task: allows to define the name of a new annotation scheme
- Rename Selected Task: changes the name of an annotation task
- Delete Selected Task: deletes an annotation task. BE CAREFUL: if you delete a task all the annotations made following that task are automatically deleted.
- Export Selected Task: exports a selected XML annotation scheme from CAT
- Import Task: import an XML annotation scheme in CAT

The **Markable** menu permits the following actions:

- Define/Modify Markable: opens up a window in which you can define the markables to be used during the annotation in a specific annotation task. For each markable it is possible to choose a label and an annotation highlight color; attributes and possible values can be also defined and modified in this window.
- Rename Selected Markable: changes the name of a specific markable.
- Delete Selection: cancels the highlighted portion of text;
- Add Markable: confirms that the highlighted portion of text is to be marked up.

The **Relation** menu permits the following action:

- Define/Modify Relation: opens up a window in which you can define relations among markables to be used during the annotation in a specific annotation task. For each relation it is possible to choose a label and an annotation highlight color; relation types, source and target types, directionality, attributes and possible values can be also defined and modified.

The **Statistics** menu permits the following action:

- Count markables in this file, as shown in the image below

atistics are computed on	ly on saved markables	
Cairo_25-November-20	12-Manuela	
Markable	# Occurences	
C-SIGNAL:	1	
ENTITY:	17	
ENTITY_MENTION:	27	
EVENT:	12	
EVENT_MENTION:	17	
SIGNAL:	8	
TIMEX3:	5	
VALUE:	4	

Figure 2: window with the number of annotated markables in a file

- Distribution of attributes on this file
- Count markables on corpus
- Distribution of attributes on corpus
- Inter-Coder Agreement: at the moment it works only on the extent of annotated markables using the Dice coefficient metric and only between

two corpora (i.e. between the annotation performed by two annotators). For more details see Section 4.

The **Help** menu contains some information about the tool (such as the license agreement and the reference to be cited in scientific papers) and the user manual in pdf format.

On the top of the Text Panel, the text panel menu bar is displayed (see figure 2).

✔ Confirm Markable	X Delete Selection	I	Ē	I	Task Selection NewsReader	/	C-SIGNAL	•	L	*	Q	Q

Figure 3: the Text Panel menu bar

The bar offers specific actions for text annotation and manipulation:

- Confirm Markable : confirms that the selected portion of text is to be marked up;
- Delete Selection : cancels the selection of a portion of text;
- Copy and paste text (important when adding the head of entity mentions): select a portion text, click on the icon and a window appears, ctrl+c to copy in the clipboard, close the window and then ctrl+v to paste into a textbox attribute.
- Task selection: drop down menu from which you can choose the annotation scheme
- /: lists the set of available markables for the selected task
- 💉 modify text panel width reducing its width
- Q Q increases and decreases text font size in the Text Panel menu

The Relation Panel (fig. 3) can be visualized pressing F1 or clicking on the dark gray bar at the bottom of the page. When activated, the Relations Panel is displayed under the Text Panel.

Through this panel it is possible to add, delete or modify relations for every relation type in a task, add, modify and delete source and target for each relation.

ALINK ‡	Relations	
9	→ CLINK	
	▼ TLINK	
	S3 searching -> TLINK -> S3 hunting	
	S3 hunting -> TLINK -> S2 gone	
	S4 asleep -> TLINK -> S3 searching	
	S5 awake -> TLINK -> S4 asleep	

Figure 4: the Relation panel

The Layer Panel has 2 tabs (fig. 4):

- Markables: each type of markable annotated in the text and displayed in the Text Panel has a checkbox which enables and disables the visibility of that type in the Text Panel. All annotations are visible by default.
- Empty tags: lists the empty tags, that is tags with no textual content



Figure 5: the Layer Panel



Please note that the default view of CAT can be partially changed:

- hide/unhide the Corpus Panel clicking on the <sup>(1)</sup> icon or moving up the dark gray splitter bar above the Layer Panel



Figure 7: hide/unhide the Corpus Panel

- hide/unhide the Relation Panel pressing F1 or clicking on the dark gray bar at the bottom of the web page

Markabies Empty Tags C-SIGNAL ENTITY ENTITY_MENTION EVENT EVENT SIGNAL TIMEX3 VALUE	S4 sparking violence in several major cities. The powerful Muslim Brotherhood has called nationwide demonstrations in support of Morsi's decree, which has put him on a collision course with the judiciary and with many of the political forces that brought down Hosni Mubarak in a 2011 popular uprising.

Figure 8: hide/unhide the Relation Panel

- modify the height of the Relation Panel when it is open moving up and down the dark gray bar above the panel
- reduce the width of the Text Panel clicking on the rext icon in the Text Panel menu bar. This option allows visualizing the Text Panel and the pop-up attribute window one next to the other.

File Edit Task Markable	Relation Statistics Help     ✓ Confirm Markable X Delete Selection	cction NewsReader 💌 / C.SIGNAL 💌   🖍 Q Q
agreement-FBK agreement-VUA Generation agreement-VUA DEMO Giscussion-FBU Giscussion-FBU Giscussion-FBK Cairo_25-November-2012 Cairo_25-November-2012 Egypt_stocks_plunge-25	Markable Attributes     X       Current     %       Extent:     Markable       VALUE     type:       type:     PERCENT       comment:	Cairo_25-November-2012-Manuela         S0       Cairo Shares Plunge 9% 33 Political Crisis Deepens.         Share prices on the Egypt Exchange declined almost 9.5 percent (s)         midday (1000 GMT) on Sunday, after President Mohamed Morsi         S1         assumed sweeping powers that sparked clashes and polarized the country's politics.         The main EGX-30 index shed 9.49 percent to reach 4,923.19 points, according to the Egyptian Exchange.         The stock exchange suspended trading for half an hour after         S3 intense selling by investors, a source at the exchange told Agence France Presse.         The drop comes as Egyptians have been deeply split by Morsi's
	CLINK         ▼         Relations           ▶         CLINK         ▼           ▶         TLINK         ▼           ▼         SLINK         SLINK           S3 told <>> SLINK         S3 stold <>> SLINK           ▶         HAS_PARTICIPANT         ▼           ▶         REFERS TO         T	uspended

Figure 9: working environment with a reduced width of the Text Panel

### 3.2 Creating a Corpus

- 1. From the top Menu, select *File*  $\rightarrow$  *New Corpus*. A pop-up window appears.
- 2. Type in the name of the corpus.
- 3. A new folder appears in the Corpus Panel, on the left-hand side of the interface.

To rename a corpus, right click on the corresponding corpus folder.

### 3.3 Adding Files to a Corpus

There are 2 options:

- 1) Import a plain text files, encoded in UTF-8 on which CAT performs a simple tokenization:
  - Click on the folder corresponding to the Corpus you have created in order to select it
  - From the top Menu, select  $\rightarrow$  *File*  $\rightarrow$  *Import*  $\rightarrow$  *Text Files*
  - Select one or more files and click *Open*
- 2) Add a tokenized file with a token per line format:
  - Click on the folder corresponding to the Corpus you have created in order to select it
  - From the top Menu, select  $\rightarrow$  *File*  $\rightarrow$  *Import*  $\rightarrow$  Tokenized Text Files
  - Select one or more files and click "Open"

The imported files are listed in the selected folder.

In order to delete or rename an imported file, *right-click* on it.

Please note that it's not possible to import a file without having previously created a corpus.

### 3.4 Importing the NewsReader Annotation Task

The already defined annotation task for the project can be imported in CAT through the following steps:

1. From the top menu select *Task*  $\rightarrow$  *Import Task* 



Figure 10: importing a task

2. Browse in your folder and select the file named CAT\_task\_NewsReader.xml

The imported task will be shown in the *Task Selection* drop-down box in the Text Panel menu bar.

✔ Confirm Markable	X Delete Selection	Î I	Task Selection	NewsReader 💌	1	C-SIGNAL	- 1	*	Ð	Q
		_		Dirha FrameNet						
				NewsReader						
				NewsReader4 prova Terence_1.0 Workshop						

Figure 11: the Task Selection drop-down box

### 3.5 Modifying an Existing Markable

In order to modify an existing markable in the NewsReader task, CAT requires the following steps:

- 1. From the top menu select *Markable* → *Define/Modify Markable* or press Ctrl-D
- 2. A new window appears where you can load the existing markable you want to modify selecting the corresponding task (i.e. NewsReader) and markable and clicking the reload icon on the right <sup>(1)</sup>
- 3. Colors and attributes related to the selected markable appears and can be modified right-clicking on them. Three options are available: Delete Attribute, Modify Attribute and Rename Attribute.

	Reader 💌 TIMEX3 💌 🕻
olor: 🛃	8
type:	DATE
value:	
functionInDocument:	NONE
anchorTimeID:	drag markable here to link
beginPoint:	drag markable here to link
endPoint:	drag markable here to link
comment:	Delete Attribute Modify Attribute

Figure 12: the Define Markable window

The order of the attributes and the order of the values in a can be changed using a drag and drop.

elect Task: NewsRead	ier 💌 🕅	4EX3		
olor:				
naa 🚛 🚛 🛌 💩 Modify Attribute			×	
Please, modify attribu	ite:			
Attribute Name: type	el			
Attribute Type: Co	mbo Box	•		Save
Values:		Q		
DATE		:		
TIME			-	
DURATION		-		
SET		2		

Figure 13: change the order of the attribute values

To add an attribute click on one of the five possible types of controls, namely:

- text box element
- radio button element
- check box element
- dropdown/combobox menu element
- reference link element



Figure 14: icons of the different types of attributes

At the moment, it is possible to rename markables from a task but not to delete them from an annotation task.

### 3.6 Modifying an Existing Relation

In order to modify an existing relation, CAT requires the following steps:

- 1. From the top menu select *Relation*  $\rightarrow$  *Define/Modify Relation* or press Ctrl-R
- 2. A new window appears where you can load the existing relation you want to modify selecting the corresponding task (i.e. NewsReader) and relation and clicking the reload icon on the right
- 3. The type of relation can be modified, from the *Relation Type* drop-down box. The options are: *one to one, one to many, one to many, many to one, many to many*
- 4. The options *Typified* (if the relation always involves a certain kind of markable) and *Directional* (if the relation holds in only one direction) can be changed as well.
- 5. Colors and attributes related to the selected relation appear and can be modified right-clicking on them. Three options are available: Delete Attribute, Modify Attribute and Rename Attribute

To add an attribute click on one of the five possible types of controls, namely:

- text box element
- radio button element
- check box element
- dropdown/combobox menu element
- reference link element

At the moment, it is not possible to rename or delete relations from a task.

### 3.6.1 Modifying Markable and Relation Attributes

After right-clicking on an attribute and choosing the option Modify Attribute, a pop-up window appears in which it is possible to:

• change attribute name

• change the type of control

In a multi value type of control (i.e. Radio Button or Combo Box) you can delete and rename each value right-clicking on it or add new values clicking on the icon  $\Box$ 

Please note that if you change from a multi value type of control to a text box control the annotated values are maintained. Vice versa, if you change from a text box control to a multi value type of control, to maintain the previously annotated values you have to add that value as a value of the new multi value attribute.

### 3.7 Opening a File

To open a file click on its name in the file list displayed in the Corpus Panel: the text will appear in the Text Panel on the right.

### 3.8 Performing Markable Annotation

3.8.1 Starting the annotation

- 1. Click on a corpus folder in the Corpus Panel on the left-hand side;
- 2. Click on the name of the file you want to annotate
- 3. The text will be displayed in the Text Panel on the right.

3.8.2 Annotating a markable

- 1. Select the NewsReader annotation task from the *Task Selection* dropdown box in the Text Panel menu bar
- 2. Select the type of markable you want to annotate from the *Markable* dropdown box in the Text Panel menu bar, containing all the defined markables for the selected task.
- 3. Select the portion of text (i.e. one or more tokens) to be annotated by pressing **shift** and clicking on the relevant token/s. To select non-contiguous tokens (as shown in the figure below), press **shift+first token** and **ctrl+shift+other tokens** (on Mac: shift**+ first token** and **CMD+shift+other tokens**).
- 4. Confirm the annotation by pressing ctrl+a or clicking the icon
   ✓ <sup>Confirm Markable</sup> in the Text Panel menu bar. Please note that the shortcut ctrl+a works in Chrome and Safari but not in Firefox.

In order to see the options available on a markable (e.g. deletion), right-click on it.

Options available on a markable are as follows:

- Delete markable
- Change Markable Extent: you can modify the extent of a markable right clicking on an already annotated markable and choosing this option. The markable becomes highlighted with a dotted line; select the new extent and ctrl+a or "confirm markable" to change the extent. If you change your

- Modify token word: you can edit the word on which you right clicked
- Define sub-tokens: you can annotate sub-sections of a token. Please note that this feature is not fully implemented yet: sub-sections of a token cannot take part in a relation.
- Involved in: lists all the relations having the markable as a source or a target. If you click on one of the listed relations, the source and the target become highlighted and the window with relation attributes opens.

To remove the token selection click on the Delete Selection icon in the Text Panel menu bar or press **ctrl + z** (this shortcut works in Chrome, Safari and Firefox).

Following the same method, it's possible to annotate overlapping portions of text. Overlapping annotated portions of text are highlighted in bold.

# Note that a file is automatically saved after every action performed by the annotator.

3.8.3 Annotating attributes of a markable

- 1. Clicking on a markable, the *Markable Attributes* pop-up window appears.
- 2. Fill in the form with correct values following the annotation guidelines.
- 3. Click "save". Note that unsaved values are shown in <u>red</u> (as specified in the caption "\*unsaved values"), e.g.:

Current Extent:	Spain
Markable type:	ENTITY_MENTION
head:	
syntactic_type:	NOM 💌
comment:	

### Figure 15: example of an attribute window with unsaved values (in red)

If default values are present they are displayed in the *Markable Attributes* pop-up window. Please note that also the default values must be saved. After saving, the attributes turn black, e.g.:

Current Extent:	Spain
Markable type:	ENTITY_MENTION
head:	Spain
syntactic_type:	NAM 💌
comment:	

Figure 16: example of an attribute window with saved values (in black)

4. To assign the value to a reference link attribute, drag and drop the corresponding markable that is to be used as a value for that attribute. To delete the value of a link attribute right click on the name of the attribute.

Current Extent:	Nulla conseq	uat massa quis enim	
Markable type:	Example_Ma	rkable	
Example_Tex_At	tribute:	Def_value	r
Example_Radio_/	Attribute:	First Second Third	p
Example_check_	Attribute:		
Example_dropdo	wn_Attribute:	Second ‡	s
Example_referen	ce_Attribute:	drag markable Markab	e Grab

Figure 17: assigning the value to a reference link attribute

5. The *Markable Attributes* window does not close automatically after saving. When finished, close the window clicking on the **Section** icon.

### 3.8.4 Modifying/deleting an annotation

To modify an annotation, click on a token belonging to the relevant markable. If the token is part of more than one markable, a window appears, asking to choose the markable you are interested in, e.g.:



Figure 18: choose the markable to modify

Select the relevant markable and confirm the selection with OK. A window appears showing the attributes, whose values can be modified.

To delete an annotation, right-click on the relevant markable, and select the option *Delete Markable*. If the token is part of more than one markable, a window appears, asking to choose the markable you want to remove. If a markable is involved in a relation, i.e. is the source or a target of a relation, a warning message appears asking you to delete the relation first.

### 3.9 Annotating a Relation

In order to set a relation:

- 1. Visualize the *Relation* panel by pressing F1 (or clicking on the dark gray bar at the bottom of the page)
- 2. The *Relation* panel shows up
- 3. Select the relation type you want to annotate from the drop down menu shown in the Relation Panel



Figure 19: choose the type of relation to annotate

*4.* Click on the **set** icon to set a new relation. A bar will appear, showing an empty relation:

	Relations
0	▼ CLINK
	Empty_CLINK_Markable



- 5. Press alt+click on the first markable (Control+Alt+Click on Linux) which is the source of the relation (i.e. the left): the markable is highlighted in the Text Panel. If the relation requires more than one source element, press alt+click on all markables to be annotated as sources.
- 6. Press alt+shift on the second markable (Control+Alt+Shift+Click on Linux) which is the target of the relation (i.e. the right argument): the markable is highlighted in the Text Panel. If the relation requires more than one target element, press alt+shift on all markables to be annotated as targets.

CLINK	•	Relations
9		▼ CLINK
		S4 move <-> CLINK <-> S4 split

Figure 21: add source and target of a relation

In case of one to one relations, you can create a new relation dragging and dropping the source markable over the target markable.

54	The drop comes as Egyptians have been deeply split by Morsi 's move to place his decisions beyond judicial scrutiny with rival rallies
	sparking violence in several major cities .

Figure 22: drag and drop the source to the target of the relation

7. The relation is shown in the *Relation* panel: when selected it is highlighted in dark grey and the corresponding source/s and target/s is/are highlighted in the text.

The drop comes as Egypti S4	ans have been deeply split by Morsi 's move to place his decisions beyond judicial scrutiny with rival rallies
	Relations
0	▼ CLINK S4 move <-> CLINK <-> S4 split

Figure 23: source and target highlighted in the text

8. In order to quickly find and highlight the argument of a relation within a text, click on the magnifier icon. Once you click on that icon, the text scrolls to the point in which the argument is located and the relation is highlighted both in the Relation Panel and in the Text Panel; in addition the background of the sentence containing the relation blinks in yellow.

3.9.1 Changing or Deleting the Source or the Target of a Relation

There are two ways to change the source or the target of a relation:

- 1. select the relation and choose different markables with alt+click and alt+shift
- 2. select the relation and right click on it: choose the "Remove source" or the "Remove target" command from the menu and choose different markables with alt+click and alt+shift. In the case the relation has more than one source or more than one target, only this second option is available.

To delete only one Source or one Target in a many-to-one, one-to-many or manyto-many relation right click on the relation and choose the element to delete from the list of sources/targets.



Figure 24: delete one source or one target in a relation

#### 3.9.2 Deleting a Relation

To delete a relation:

- 1. select the relation to be deleted
- 2. click on the icon in the Relations Panel or right click on the relation and select the "Delete Relation" option.

### 3.9.3 Annotating Relation Attributes

To annotate the attributes related to a relation:

- 1. double-click on the relation in the Relation Panel in order to open up the attributes window
- 2. fill in the form with correct values. To assign the value to a reference link attribute, drag and drop the corresponding markable that is to be used as a value for that attribute. To delete the value of a link attribute right click on the name of the attribute.

3. click on the "Save" icon on the bottom right of the attributes window.

Note that unsaved values are shown in <u>red</u> (as specified in the caption "\*unsaved values"). If default values are present they are displayed in the *Markable Attributes* pop-up window. Please note that also the default values must be saved.

After saving, the attributes turn black.

### **3.10 Annotating Empty Tags**

CAT allows to annotate empty tags, that is tags with no textual content, for every type of markable defined for a task. For the NewsReader project two markables are defined as empty tags, namely <ENTITY> and <EVENT>.

- 1. Select the NewsReader annotation task from the *Task Selection* dropdown box in the Text Panel menu bar
- 2. Select the type of markable you want to annotate from the *Markable* dropdown box in the Text Panel menu bar, containing all the defined markables for the selected task.
- 3. Go to the Layer Panel below the Corpus Panel on the left side and choose the "Empty Tags" tab.
- 4. Click on the sicon to set a new empty tag for the selected type of markable: all the empty tags have the same default name "Empty\_Mark". To change the name of an empty tag, right click on it and select the "Rename" option and follow the instructions. This corresponds to the tag\_descriptor attribute of the <ENTITY> and <EVENT> markables. Names of empty markables should follow the rules defined in the annotation guidelines of the project.

21/24



Figure 25: empty tags

- 5. To annotate attributes double click on the empty tag: the attributes window opens up. Fill in the form with correct values. Please note that to assign the value to a reference link attribute, drag and drop the corresponding markable that is to be used as a value for that attribute. To delete the value of a link attribute right click on the name of the attribute. Click on the "Save" icon on the bottom right of the attributes window.
- 6. Note that unsaved values are shown in red (as specified in the caption "\*unsaved values"). If default values are present they are displayed in the Markable Attributes pop-up window. Please note that also the default values must be saved. After saving, the attributes turn black. To delete an empty tag, right-click on it and select the option "Delete".

An empty tag can be involved in a relation by adding it as source or target.

### **3.11 Hiding Annotated Markables**

Annotations may be hidden from view using the Layers Panel. Each type of markable has a checkbox in the "Markables" tab of the Layers Panel which enables and disables the visibility of the annotation in the Text Panel. All annotations are visible by default.

### 3.12 Saving a File

A file is automatically saved after every action performed by the annotator.

### **3.13 Exporting a Corpus**

To export the annotated files of a corpus:

- 1. Select the corpus folder you want to export in the Corpus Panel
- 2. Select the task NewsReader from the *Task Selection* drop-down box in the Text Panel menu bar

3. Select *File*  $\rightarrow$  *Export as stand-off XML* and wait until the zip folder is downloaded

Note that ALL the files of the selected corpus (both annotated and non-annotated) will be exported.

The exported corpus can be imported again in CAT:

- 1. Create a new corpus in CAT as explained in section 3.2
- 2. Select *File*  $\rightarrow$  *Import XML CAT file* and upload one or more files.

## 4. Calculating Inter-Coder Agreement

CAT contains a built-in functionality for the calculation of the Inter-Coder Agreement on two corpora. At the moment it works only on the extent of annotated markables using the Dice coefficient.

The 2 corpora should have different names but contain files with the same names: in case the two corpora have files with different names, the script calculates the agreement only on those files with the same name.

Procedure:

- 1. Select Statistics  $\rightarrow$  Inter-Coder Agreement
- 2. A window pops us: drag and drop the two corpora in the window
- 3. Different measures appear in the window, namely:
- Overall results: Micro and Macro average on all the files with the same name
  - Dice coefficient on markable extent for each type, perfect overlap
  - Dice coefficient on markable extent for each type considering only the overlapping tokens
- Results for each single file
  - Dice coefficient on markable extent for each type considering a perfect overlap
  - Dice coefficient on markable extent for each type considering only the overlapping tokens

## 5. List of Keyboard Shortcuts in Alphabetical Order

- Confirm markable: Ctrl+A (not working in Firefox)
- Define a new markable: Ctrl+D (minor issue in Firefox)
- Define a new relation: Ctrl+R (minor issue in Firefox)
- Delete selection: Ctrl+Z
- Find in a text using the browser search feature: Ctrl+F
- Open up the Relation Panel: F1
- Select portion of text to be annotated: Shift+Click/Shift on text

- Select non-contiguous tokens: press shift+first token and ctrl+shift+other tokens (on Mac: shift+ first token and CMD+shift+other tokens)
- Select the source/s of a relation in Windows and Mac: Alt+Click
- Select the target/s of a relation in Windows and Mac: Alt+Shift+Click
- Select the source/s of a relation in Linux: Control+Alt+Click
- Select the target/s of a relation in Linux: Control+Alt+Shift+Click

## 6. Updates from the previous version of the manual

- Correction of the shortcut to be used to annotate non-contiguous tokens.

### References

Lenzi, V. B., Moretti, G., & Sprugnoli, R. (2012). CAT: the CELCT Annotation Tool. In LREC (pp. 333-338).