SACR: User Guide

Bruno Oberle

2018-09-09 (v0.0)

Contents

1	Introduction 2							
	1.1	Presentation	2					
	1.2	Important terminology	3					
2	The	nterface at a glance	4					
	2.1	Main window	4					
	2.2	Popup	5					
3	Reco	nmended workflow	6					
4	Crea	ing, saving and opening a new document	8					
	4.1	How to import or open a text with or without annotations	8					
	4.2		10					
	4.3		10					
	4.4		l 1					
5	Creating and editing annotations 12							
	5.1	Creating mentions	12					
	5.2		13					
			13					
			۱4					
			۱4					
			15					
			15					
			16					

		5.2.7 Using ctrl (batch mode)	16			
	5.3	Selecting a mention	16			
	5.4	Detaching an element from a group	17			
	5.5	Merging two sets	17			
	5.6	(Re)naming a set	17			
	5.7	Changing the color of a chain	19			
5	Sear	Searching and visualizing				
	6.1	Displaying only one set	19			
7	The	popup	19			
	7.1	Command reference and all other commands that aren't listed elsewhere in this				
		guide	19			
3	Sett	Setting mention properties				
	8.1	Schema format	23			
	8.2	How to change the schema after the annotation is started	25			
	8.3	Slow mode	25			
	8.4	Quick mode	25			
9	Misc					
	0 1	How to convert from and to the Clozz format	25			

1 Introduction

1.1 Presentation

SACR is an easy to use annotation tool, specifically designed to annotate coreference chain. It is written in HTML, CSS and Javascript, and consists in one web page. You can use it freely at http://boberle.com/projects/sacr or download it from there. It is distributed under the terms of the Mozilla Public Licence, version 2.0 (see de LICENSE file).

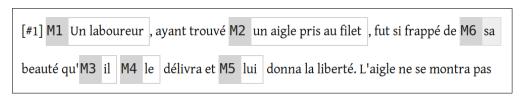
If you use it, please cite:

Oberlé B. (2018). SACR: A Drag-and-Drop Based Tool for Coreference Annotation. Proceedings of the 11^{th} Edition of the Language Resources and Evaluation Conference (LREC 2018). Miyazaki, Japan.

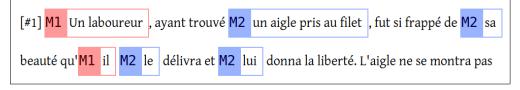
```
@inproceedings{Oberle-2018-sacr,
   author = {Bruno Oberle},
   title = "{SACR: A Drag-and-Drop Based Tool for Coreference Annotation}",
   booktitle = {Proceedings of the Eleventh International Conference on
        Language Resources and Evaluation (LREC 2018)},
   year = {2018},
   month = {May 7-12, 2018},
   address = {Miyazaki, Japan},
   editor = {Nicoletta Calzolari},
   publisher = {European Language Resources Association (ELRA)},
   isbn = {979-10-95546-00-9},
   language = {english}
}
```

Remember to use Firefox (or at least Chromium or Google Chrome, but definitevely not Internet Explorer or Edge).

With SACR, you create mentions:



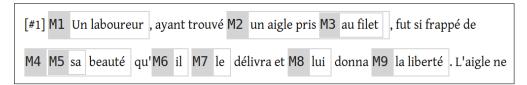
and you group them into sets:



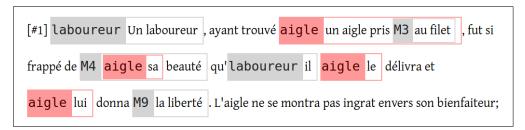
SACR is designed to annotate coreference, so usually "mentions" are *referring expressions* and "sets" are group of referring expressions that refers to the same referent. These sets are sometimes called *coreference chain*. But you can use it to annotate something else.

1.2 Important terminology

We will deal with mentions:



Mentions can be grouped in a set (note that the name is the same for each member of the set):



When the number of mentions of a set reaches a certain threshold (usually 2 or 3, depending of the task), the are colored. They will then be called (*coreference*) *chain*. In the previous illustration, for example, the threshold is set to 3: you see that there are five *sets*: "laboureur", "aigle", "M3", "M4", "M9", but also that there is only on *chain*: "aigle" since only this set has three elements or more.

So a *mention* is an element of a *set*. A set big enough to reach a certain threshold is called a *chain*. Note that all *chains* are *set*, but that the reverse is not true.

2 The interface at a glance

2.1 Main window

The figure 1 illustrates the main interface. It is composed of the following elements:

- 1. Metadata or comments in the source file.
- 2. An unselected marked referring expression part of the chain "laboureur".
- 3. A selected marked referring expression part of the chain "aigle".
- 4. An unselected marked referring expression ("au filet") which isn't part of any chain (so it is gray).
- 5. A set of features to be annotated, like the POS of the syntactic head of the referring expression.
- 6. List of possible values for a feature. You just have to select the correct one.

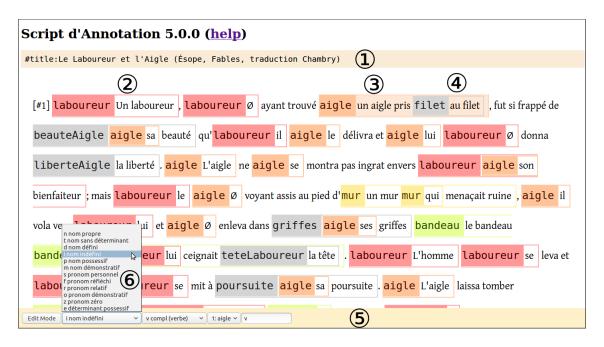


Figure 1: The main interface.

As you can see, there is no control buttons. Commands are activated by clicking on keys on your keyboard. For example, to change the color of a chain, you need to press c (as in "Color").

To get a list of all available commands, press h.

2.2 Popup

The figure 2 illustrates the popup window that lists all the chains and the elements:

- 1. The main interface in the background.
- 2. The chain "laboureur" (the list of its expressions has been collapsed).
- 3. The chain "aigle" with the list of its expressions.
- 4. The expression "un aigle pris au filet" is selected both in the main window and in the list. You can select in one or the other. Note that the chain name "aigle" is surrounded by a black frame.
- 5. Other chains with their expressions.

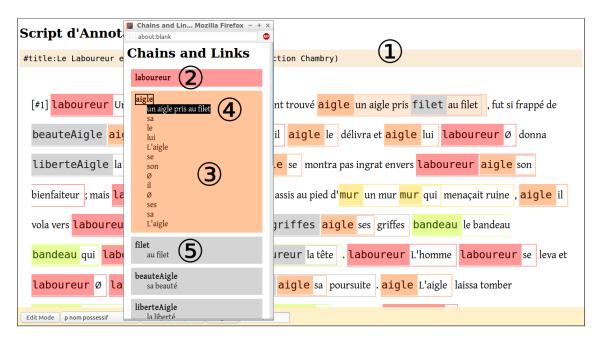
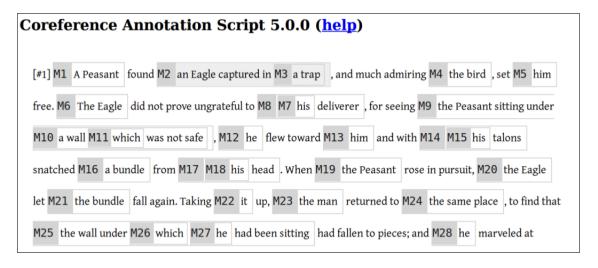


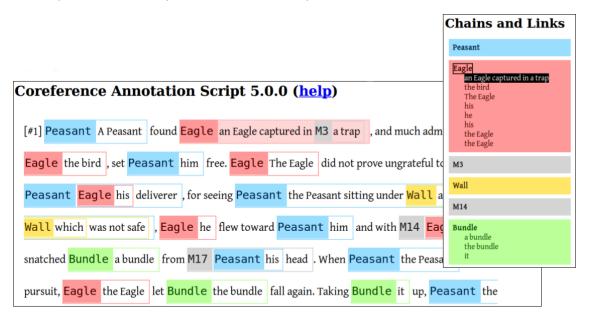
Figure 2: List of all chains and elements in a popup window.

3 Recommended workflow

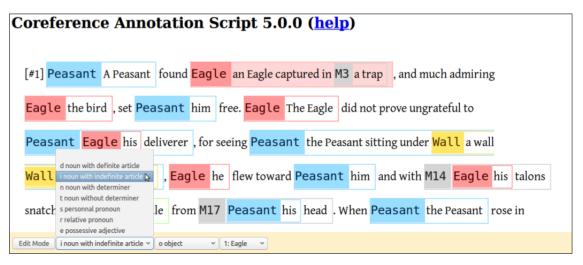
First, create the mentions (referring expressions):



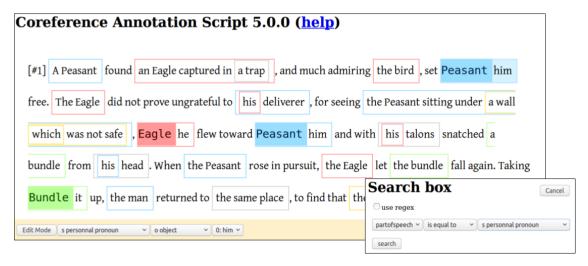
Second, create the sets (coreference relations):



Third, annotated features for each mentions, if needed:



Fourth, visualize your data:



4 Creating, saving and opening a new document

4.1 How to import or open a text with or without annotations

When you first open the web page, the tool will ask you to load a text and, if you want, an annotation scheme (or schema), that is, the list of the properties (and their possible values) you want to annotate for each referring expression.

To load a text with the data loader (figure 3):

- 1. Copy a text (from the Web, Writer, Word, etc.) and copy it into the box, or use the button *Browse* to import the text from a file stored on your computer. You may also use a sample text (an Aesop's fable) with or without sample annotations by clicking on the buttons.
- 2. If you want to annotate a set of features for each mentions, you may copy a schema in the next text box, or import it from a file stored on your computer (using the button *Browse*), or use a default sample schema (to work with the sample texts).
- **3.-6.** Set up the configuration options (see section 4.2).
- **7.** Finally, just click on the parse the data button.

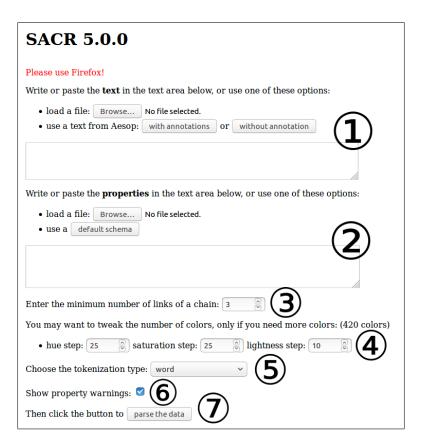


Figure 3: Data loader.

4.2 Configuration

See figure 3.

Enter the minimum number of links of a chain: set when a set is big enough to be a chain and receive it's own color. Usually 2 or 3, depending of the task and the size of your text.

You may want to tweak the number of colors, if you need more colors. By default, there are 420 colors. The more color you add, the less distinctive they will be. Play with the hue, saturation and lightness to increase or decrease the number of colors.

Choose the tokenization type:

- word for normal text with latin alphabet. Punctuation *cannot* be selected and be a mention on their own.
- word for normal text with latin alphabet. Punctuation *can* be selected and be a mention on their own.
- character: every character can be selected individually. Use this for language such as chinese or korean.

Show property warnings: if you check this box, SACR will report an error if there is a difference between the text and the schema you imported. For example, if there is a property category in the schema, but not in the text.

4.3 Text and annotation format

Annotations are inline, that means that there are stored within the text. Metadata are also stored within the text.

The text must be in a plain text format. Paragraph are separated by white lines. Two consecutive non white lines are considered as been in the same paragraph.

Here is an example:

```
This is a sentence of the first paragraph... which continues here... and even here!

After the white line, this starts the second paragraph.
```

Metadata are added as lines that start with #. Some metadata are used by SACR itself, as #COLOR. While #textid and #title are not mandatory, they are exploited by SACR to display the text id and the title in the browser window title bar. Other metadata may be used to organize the corpus (as #textid, #title, #source, etc.) or by the tool that performs the statistical analysis (for example, to distinguish between parts (introduction, body, conclusion, etc.):

```
#textid: 1
#title: text title
This is the text...
```

Two metadata are important:

- Lines starting with #additionnaltoken: new token are used to define additionnal tokens. For example if you want a symbol (like "α") to be recognized as chunk of text that you can annotate, you can use this metadata.
- Lines starting with #part-heading:level=1 indicate that the following paragraph is a heading. This is used for a better layout.

Annotations are stored as follows:

```
{ReferentName:prop1="val1",prop2="val2" text inside the mention}
```

For example:

```
{chat:categorie="SN défini",fonction="sujet" Le petit chat de {voisine:categorie="SN défini",fonction="cdn" la voisine}}
```

See figure 4 for a complete example.

4.4 Saving the annotations and opening them later

Before leaving the page, you need to save your annotations in a file. Just press w (as in "Write"). Depending on your browser settings, the file will be saved automatically (usually in your "Downloads" directory) or you will be asked where to store it.

The filename will include a time stamp (that is, the current date and time). This allows you to save several versions of the same file.

Imagine you have loaded a file called esope. When you press the w key, SACR will save a file called esope_20170529_230130 (where 20170529_230130 is the date (year, month, day) and time (hours, minutes, seconds).

If you save the some minutes later, you get, for example, esope_20170529_230424 (the name ends with 0424 instead of 0130).

Figure 4: Text example.

If you copied a text in the text area without using the *browse* button on the data loader, you get a name like default_20170529_230130.

The next annotation session, you just load (using the "browse" button of the data loader) the last file you saved, for example esope_20170529_230424 or default_20170529_230130.

5 Creating and editing annotations

5.1 Creating mentions

To create singletons:

- if the mention have only one word: double click on that word,
- if the mention has several words:
 - click on the first word (it is then underlined; to cancel and choose another word, just click again on it and it will be un-underlined),
 - click on the last word.

Batch mode:

- select a mention,
- create mention as before, but hold ctrl: all the newly created mentions will be automatically attached to the mention you selected in the first step.

Example of the batch mode:

- Let's say you have the sentences: "Paul est heureux. Il est content. Il est joyeux."
- Create a mention for "Paul" and then select it.
- Hold ctrl and double-click on the first "il" and the second "il".
- Both "il" are attached to "Paul": you have create the set "Paul... Il..." with a minimal effort!

5.2 Creating sets (coreference relations)

Let's say you have several mentions, and you want to group them (for example because they are coreferential).

5.2.1 Using drag-and-drop

You just need to drag-and-drop one mention one over another. The mention that is attached to the set will get the name of the set. If the set reach the critical size to be chain (see the terminology in section 1.2), then it will be colored.

There are some behavior differences depending of the source and the destination, but the process is designed to be intuitive. Here is the detail (note that a "singleton" is a set with one element, and a "group", in the table, a set with more than one element):

source	destination	result
singleton	singleton	destination is attached to the source
singleton	group	source is attached to the destination (1)
group	singleton	destination is attached to the source
group	group	groups are merged (SACR will ask you to confirm) (2)

Notes:

• (1) If you want to detach the destination from its group and attached to the source, hold ctrl.

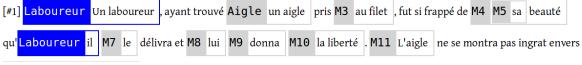
• (2) If you don't want to merge, but just detach the destination from its group and attached it to the source, hold ctrl.

It's a bit abstract, but in fact it's intuitive. Just practice!

5.2.2 Example 1: singleton to singleton

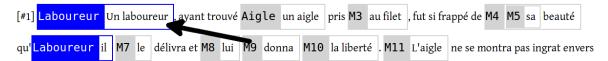
Drag the singleton *Un laboureur* and drop it over the singleton *il*:





5.2.3 Example 2: singleton to group

Drag the singleton *donna* and drop it over one of the mentions of the group, here for example *Un laboureur*:



The mention *donna* is attached to the group:



5.2.4 Example 3: group to singleton

Do the opposite: drag one of the mentions of the group, here for example *Un laboureur* and drop it over the singleton *donna*:



You get the same result as before:



5.2.5 Example 4: group to group

Drag one of the mentions of a group, here for example *Un laboureur*, and drop it over one of the mentions of another group, here for example *lui*:



You will be ask if you want to merge:



If yes:



5.2.6 Example 5: Using ctrl

If you hold ctrl, you detach the destination mention (here un aigle) from its group:



to attach it to the source group:



5.2.7 Using ctrl (batch mode)

- Select a mention,
- hold ctrl and click on other mentions: they will all be attached to the mention you selected in the first step.

Example of the batch mode:

- Let's say you have the sentences: "Paul est heureux. Il est content. Il est joyeux."
- Create a mention for "Paul", "Il" and "Il".
- · Select "Paul".
- Hold ctrl and click on the other two mentions.
- Now, they all belong to the same group!

You can use this if you have a to relate to distant mentions. Select the first, scroll down to the second one and click on it.

5.3 Selecting a mention

Don't confuse "selecting (or clicking on) a mention" and "selecting (or clicking on) a word". You select a word by clicking on it: it is underlined. You do that to create a new mention (you need to click on another word to complete the process, see section 5.1). If you have selected a word by error, just click on it a second time to deselect it.

To "select (or click on) a mention", you need to click on its name (the part with the colored background) or anywhere on the white part inside the frame of the mention, except on a word (in which case you select a word...). To deselect a mention, click on it.

See figure 5 for some examples.

When you select a word, all mentions are deselected, and vice-versa.

5.4 Detaching an element from a group

Hold shift and click on the mention you want to detach: it will be grayed and renamed with a name as M123 (using the next available counter value). It is now a singleton.

5.5 Merging two sets

Drag-and-drop the element of one set over the element of another set. You will be ask if you really want to merge.

The name of the resulting set is the name of the name of the source.

See example in section 5.2.5.

5.6 (Re)naming a set

Select a mention of the set and and press:

- N to change the name of the selected chain (don't ask, use the content of the selected mention),
- n to change the name of the selected chain (ask, default is current name),
- m to change the name of the selected chain (ask, default is the content of the selected mention).

• The **word** *sa* is selected (since it is underlined): sa beauté qu'il le délivra • The **mention** *un* aigle is selected: [#1] Laboureur Un laboureur, ayant trouvé Aigle un aigle pris au filet, • The **mention** *un laboureur* is selected: [#1] Laboureur Un laboureur, ayant trouvé Aigle un aigle pris au filet, • The **mention** *qui* is selected. It is nested in another mention: au pied d'Mur un mur Mur qui menaçait ruine , il vola • The **mention** un mur qui menaçait ruine is selected. Another mention is nested (qui), which is not selected. Look at the dashed frame around un mur qui menaçait ruine (selected), and the solid frame around qui (not selected): au pied d'<mark>Mur</mark> un mur <mark>Mur</mark> qui menaçait ruine , il vola • To select a **mention**, click on the name (with the colored background) (black arrow), or anywhere in the white area (red arrow). The first method is easier: [#1] Laboureur Un laboureur

Figure 5: Word and mention selectin.

5.7 Changing the color of a chain

Reminder: a set has a color only only if it is a chain, that is if it is has as many elements as indicated in the Enter the minimum number of links of a chain textbox (see section 4.2). Other sets (not chains) are in gray.

Select a mention of the chain and press c. Choose the color and click on "cancel" (top-right corner).

6 Searching and visualizing

6.1 Displaying only one set

Select a mention of the set and press o. Press O to show all mentions.

Press u to hide all sets that are not chains.

7 The popup

TODO

7.1 Command reference and all other commands that aren't listed elsewhere in this guide

Type h to show the main available commands (see figure 6).

Creating mentions To create a mention, click on the first and the last word of the mention. If the mention has only one word, just double click on it. A default name will be used, or:

- hold shift to be asked for a name,
- hold ctrl to automatically attach the link to the last selected set.

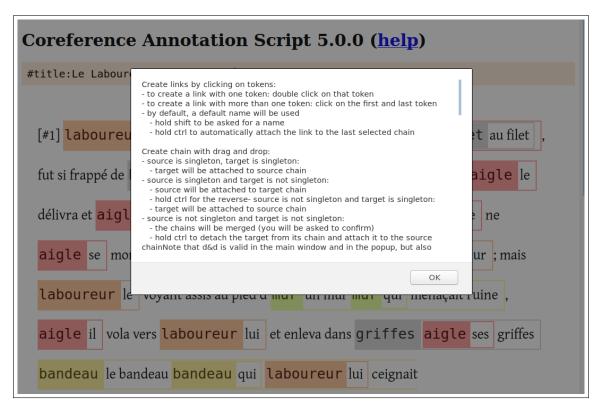


Figure 6: Help!

Creating coreference relation To attach a mention to an other (to create a "coreference relation"), drag-and-drop one mention over another.

Severals cases are possible (see below for illustrations):

- source is singleton, target is singleton:
 - target will be attached to source chain,
- source is singleton and target is not singleton:
 - source will be attached to target chain
 - hold ctrl for the reverse
- source is not singleton and target is singleton:
 - target will be attached to source chain
- source is not singleton and target is not singleton:
 - the chains will be merged (you will be asked to confirm)
 - hold ctrl to detach the target from its chain and attach it to the source chain

Note that d&d is valid in the main window and in the popup, but also between them.

Replacing a mention To replace a mention, create a new mention, then drag-and-drop, holding shift (or ctrl+shift for Firefox 54 and below), it to the mention to be replaced. This will copy all the metadata (mention name and properties) to the target mention, and remove the source mention.

Detaching a mention from its set To detach the mention, you can either:

- hold shift and click on a mention. The mention will be a singleton and its name will change to something like M123 (next available counter value),
- hold ctrl+shift and click on a mention. The mention will be a singleton but you will be ask a name (rather than M123),
- hold ctrl and click on a mention. The mention will be attached to a the last selected set.

Editing links Use:

- suppr or backspace to destroy the selected mention,
- c to change the color of the selected set,

- N to change the name of the selected set (don't ask, use the content of the selected mention),
- n to change the name of the selected set (ask, default is current name),
- m to change the name of the selected set (ask, default is the content of the selected mention).

Navigating Use:

- f to go forward in text,
- F to go forward in text (only visible),
- j to go forward in set,
- J to go forward in set (only visible),
- b to go backward in text,
- B to go backward in text (only visible),
- k to go backward in set,
- K to go backward in set (only visible),
- t to scroll to the selected mention in the main window,
- T to scroll to the selected mention in the popup window.

Showing and hiding mentions Use:

- o to show only the set of the selected mention,
- 0 to show all mentions,
- s to show the search box.
- u to hide non chains,
- U to hide/show (toggle) non chains in popup,

Popup Use:

- p to show the popup window of all sets and mentions,
- e to expand all sets in the popup,
- E to collapse all sets in the popup,
- v to expand chains in the popup.

Saving Use:

- w to write the text and annotations to a file,
- W to write the text and annotations to a file (with head text and link content),
- x to write the schema to a file,
- X to show the schema (in a dialog).

Displaying

- i to increase font size,
- I to decrease font size,
- 1 to whow statistics,
- P to show chain patterns.

8 Setting mention properties

8.1 Schema format

A schema is a list of all the properties you want to annotate for mentions, and all the their possible values. You don't need a schema if you just want to annotate coreference. But a schema is required if you want to annotate the part the speech, the grammatical function, etc. for each mention.

See figure 7 for an example which describes the main options.

You may add other parameters (in the form param1=val1, param2, val2 etc.):

- newline=true|false: add a new line characters after the property (use that if you have lot of properties),
- showname=true | false: show the name of the property, not just a text box or a list of values,
- textboxsize=4: if type is text, set the size of the textbox (4 for example),
- addShortcuts=true|false: automatically add shortcut keys for the quick mode (see section 8.4).

```
# line beginning with # are comments (i.e. there are ignored
# a new property beginning with PROP: and a serie of parameters
PROP:name=partofspeech
# the following line means "blank", no value
$$$
proper name
indefinite noun phrase
definite noun phrase
# properties are separated by a white line
PROP:name=fonction
$$$
subjet
object
. . .
# a property of type "text" allows users to put anything they want
PROP: name=comment, type=text
# the head property is special, see below
PROP: name=head, type=head
```

Figure 7: Sample schema.

8.2 How to change the schema after the annotation is started

If you want to change the schema after the annotation is started, you must load the annotations and the new schema as usual, but be careful *not* to check the Show property warnings checkbox.

SACR will automatically remove the properties that don't appear in the new schema, add the properties that don't appear in the current annotation state, and adapt the values. Of course, you will need to manually annotated the new properties and check the values if needed.

8.3 Slow mode

TODO

8.4 Quick mode

TODO

9 Misc

9.1 How to convert from and to the Glozz format

TODO