

SACR: A Drag-and-Drop Based Tool for Coreference Annotation

Bruno Oberle

LiLPa, University of Strasbourg
oberleb@unistra.fr

Context

- Statistical methods in NLP lead to an increased need for annotated corpora, but manual annotation is costly, both in terms of time and human resources. Annotation is often done by students, interns or other non-technical users.
- To speed up the annotation process, tools should be designed and optimized for a specific task. They should also be ready-to-use and provide an intuitive interface.

Presentation of SACR

- SACR (“Script d’Annotation des Chaînes de Référence”) is a cross-platform tool, implemented as a single web page, running in Firefox and Chrome.
- Easy to use, with no learning time and no installation, it is well-suited for users with no specific computing skills.
- The format is text-based, with inline annotations that can be converted from or to stand-off annotation formats like the Glozz format (Widlöcher & Mathet 2012), readable by several other tools like TXM (Heiden 2010), Analec (Landragin, Poibeau & Victorri 2012).
- It is open source (MPL-2.0) and available at boberle.com/projects/sacr.

Annotating with SACR

1. Referring Expressions

Coreference Annotation Script 5.0.0 ([help](#))

[#1] M1 A Peasant found M2 an Eagle captured in M3 a trap , and much admiring M4 the bird , set M5 him free. M6 The Eagle did not prove ungrateful to M8 M7 his deliverer , for seeing M9 the Peasant sitting under M10 a wall M11 which was not safe , M12 he flew toward M13 him and with M14 M15 his talons snatched M16 a bundle from M17 M18 his head . When M19 the Peasant rose in pursuit, M20 the Eagle let M21 the bundle fall again. Taking M22 it up, M23 the man returned to M24 the same place , to find that M25 the wall under M26 which M27 he had been sitting had fallen to pieces; and M28 he marveled at

- A click on the first and last words of a referring expression (*i.e.* a linguistic expression that refers to some extra-linguistic entity) is enough to mark it.
- Marked expressions are surrounded by a color frame.

2. Coreference

Coreference Annotation Script 5.0.0 ([help](#))

[#1] Peasant A Peasant found Eagle an Eagle captured in M3 a trap , and much admiring Eagle the bird , set Peasant him free. Eagle The Eagle did not prove ungrateful to Peasant Eagle his deliverer , for seeing Peasant the Peasant sitting under Wall a Wall which was not safe , Eagle he flew toward Peasant him and with M14 Eagle his talons snatched Bundle a bundle from M17 Peasant his head . When Peasant the Peasant rose in pursuit, Eagle the Eagle let Bundle the bundle fall again. Taking Bundle it up, Peasant the

Chains and Links

Entity	Referents
Peasant	Eagle, The Eagle
Eagle	an Eagle captured in a trap, the bird, The Eagle
Wall	the Peasant sitting under, a wall
Bundle	a bundle, the bundle, it

- Drag-and-dropping an expression over another marks a coreference link (two expressions are coreferential when they both refer to the same entity in the extra-linguistic world).
- Two coreferential expressions are assigned the same color and name.
- A pop-up window lists all the referents and their corresponding expressions. It is possible to drag and drop expressions from the main window to the pop-up and vice-versa.

3. Features

Coreference Annotation Script 5.0.0 ([help](#))

[#1] Peasant A Peasant found Eagle an Eagle captured in M3 a trap , and much admiring Eagle the bird , set Peasant him free. Eagle The Eagle did not prove ungrateful to Peasant Eagle his deliverer , for seeing Peasant the Peasant sitting under Wall a wall which was not safe , Eagle he flew toward Peasant him and with M14 Eagle his talons snatched Bundle a bundle from M17 Peasant his head . When Peasant the Peasant rose in

Wall

- d noun with definite article
- n noun with indefinite article
- n noun with determiner
- t noun without determiner
- s personal pronoun
- r relative pronoun
- e possessive adjective

Edit Mode | i noun with indefinite article | o object | 1: Eagle

- Drop-down menus are used to annotate features, like part of speech or grammatical function.
- It is also possible to use shortcut keys (*e.g.* “d” for a definite noun phrase or “r” for a relative pronoun). The program then goes automatically to the next expression to be annotated.

4. Search

Coreference Annotation Script 5.0.0 ([help](#))

[#1] A Peasant found an Eagle captured in a trap , and much admiring the bird , set Peasant him free. The Eagle did not prove ungrateful to his deliverer , for seeing the Peasant sitting under a wall which was not safe , Eagle he flew toward Peasant him and with his talons snatched a bundle from his head . When the Peasant rose in pursuit, the Eagle let the bundle fall again. Taking Bundle it up, the man returned to the same place , to find that the

Search box

use regex

partofspeech is equal to s personal pronoun

search

Edit Mode | s personal pronoun | o object | 0: him

- SACR also offers to search any features and to show the corresponding expressions in the main window.

References

- Heiden S. (2010). The TXM Platform: Building Open Source Textual Analysis Software Compatible with the TEI Encoding Scheme. In Ryo Otaguro, *et al.* (eds.), *24th Pacific Asia Conference on Language, Information and Computation*. Sendai, Japan. Wasea University.
- Landragin F., Poibeau T. and Victorri B. (2012). Analec: a new tool for the dynamic annotation of textual data. In Nicoletta Calzolari *et al.* (eds). *Proceedings of LREC'12*. Istanbul, Turkey. ELRA.
- Widlöcher A. and Mathet Y. (2012). The Glozz Platform: A corpus annotation and mining tool. In *Proceedings of the 2012 ACM Symposium on Document Engineering*, Paris, France.

Acknowledgement

This work is part of the Democrat project (DEscription et MOdélisation des Chaînes de Référence: outils pour l’Annotation de corpus (en diachronie et en langues comparées) et le Traitement automatique), supported by the French National Research Agency (ANR) (ANR-15-CE38-0008).

