ANNOTATION OF CONTENT TYPES - v1.0

1. INTRODUCTION

These guidelines discuss the annotation of content types in English texts.
In the following we present the content types to be annotated and we provide guidelines for marking them manually in texts. In order to obtain clean data, it is important to read and follow the guidelines.

Content types indicate the semantic and functional characteristics of communication in a piece of texts. Their definition is based on internal, linguistic and semantic features of texts. Their annotation is performed at clause level.
Italic is used to present examples taken from real texts.

2. ANNOTATION PROCEDURE

2.1 Before Annotation

- Read the provided abstract.

For each document or book a short summary of its content will be provided. This summary should be carefully read because it helps in identifying what is the primary narrative (see Section 3.1 for the definition of primary and secondary narratives).

- For each clause to be annotated, read the whole sentence and the following one. It's necessary to read the entire sentence containing the clause to annotate and the following sentence before starting the annotation, so to have a clear idea of what is the functional and semantic contribution of each clause to the context.

2.2 During annotation

- Each clause should be annotated with a single label. When dealing with a discourse unit whose content is ambiguous, choose the mode with the highest priority and assign that to the clause. Priority given to the labels is described in Section 4.
- Tags are assigned mainly on the basis of the information contained in the document: annotators can rely on external knowledge in case information crucial for the annotation is missed. For example, annotators can search on the web if a snippet reported within quotes is a literary reference (for more information about the annotation of literary references see Section 3.8).
- Annotation is performed at clause level thus annotators are first asked to select a clause and then to assign a label. Details on the rules to follow are detailed in the following section.

2.2.1 Clause Segmentation

A clause is a group of words related to each other and containing a subject and a verb (finite - past present future - and nonfinite - infinitive, participles, gerunds), although the subject may be implicit or be shared with other clauses. The second clause in the example below has the subject shared with the first one and contains a nonfinite verb:

> *Between these trees a peasant woman was walking, // bearing upon her head an immense brush-heap*

RULES FOR CLAUSE SEGMENTATION
- Subordinate clauses include the subordinating element introducing them
  Zelphine says // that it would be very ungrateful of us // if we were to complain of cloudy weather
- Coordinating conjunctions are included in the coordinated independent clauses
  she smiled blandly, // lowered some money in her basket, // and the pair parted with bows and compliments .
- Restrictive relative clauses, that is the ones that help to identify the referent of the word they modify, are not to be separated from the clauses they depend on
  Beside the really fine figure of the Genoese navigator is a woman who represents either Columbia or an Italianized American Indian
- Discontinuous units are required when a discourse unit of any kind interrupts the flow of another discourse unit (see the clause underlined below)

the guild of smiths, // who revere St. Eloy as their patron saint, //invoked his aid

- Non-restrictive relative clauses (and parenthetical): they form a clause of their own. In case the text segment following the non-restrictive relative clause is not a clause of its own it must be tagged as part of the clause preceding the non-restrictive clause - in the following example the portions of the text belonging to the same clausal unit are underlined
This whole morning we spent in the Aquarium ,// which is down near the sea ,// a part of the Villa Nazionale .

3. DEFINITIONS AND EXAMPLES

In this section we define the different kinds of CONTENT TYPES each corresponding to different markables. Each markable has a set of attributes: in some cases assigning a value to these attributes is mandatory while in other cases is optional.
MANDATORY
- NARRATIVE: level
- DESCRIPTIVE: type
- OTHER: type

OPTIONAL, only if the condition applies to the tagged clause
- ALL CONTENT TYPES: conversational-direct, conversational-reported

In addition, the "comment" attribute can be used by the annotators to insert comments.

3.1 NARRATIVE

Narrative clauses are characterized by the presence of events and states that can be anchored to a hypothetical timeline. Even if in the document this narration could not have a perfect sequential order, due for example to flashbacks, all the events are anchored to time often thanks to the presence of explicit temporal expressions (e.g. dates or references to week days) of temporal signals (e.g. prepositions, conjunctions or adverbs).

We left Cava on Wednesday,// and made the tour from there to Amalfi in the brilliant morning sunshine.

Of course the sun was shining // when we left Genoa .

The clause introducing a direct or a reported speech is always to be annotated as NARRATIVE.
I suggested
She says

The attribute CONVERSATIONAL should be used to specify if a clause is the content of a direct or reported speech (spoken or written text transcribed in its original form). Indeed this attribute has two values:
- DIRECT: indicates that the annotated clause is part of a direct speech.

"I never cut an orange from a tree but once in my life,// and that was in Granada ."
- REPORTED: indicates that the annotated clause is part of a reported speech.
  She was the Queen of the country,// she avowed

It is important to distinguish secondary narration from the primary one. The primary narration is the one related to the main story that is the main object of the document. All the digressions are to be considered secondary narration: they usually show a change in the main characters, in the time and sometimes also a change in the location with respect to the primary narration.

The attribute SECONDARY has only one value:
1. YES: it should be selected to label all the narrative clauses that are not primary. Leave it blank if the clauses refer to the primary narrative.

The following sentence contains clauses related to the primary narrative (in bold) and the a secondary narrative.

> ***When we were in the midst of the silvery blueness , // watching with some apprehension a small boy who dives into the water // to show off its wonderful color , // our boatman suddenly became loquacious , // and told us thrilling tales of unfortunate visitors to the grotto*** *// who had been walled in by the sea // and were obliged to spend days and even weeks in this drear abode ,// living upon supplies which daring sailors , //_who contrived to get their boats near to the mouth of the grotto_//, handed in to them .*

The factuality of the events in the clause does not influence the assignment of the category. Clauses with hypothetical, probable, uncertain, future events can be annotated as NARRATIVE.

3.2 EVALUATIVE

Evaluative clauses are those containing opinions and comments having explicit evaluations markers (in bold below):
- Comment: *some of the appointments of this **rather expensive** albergo are not equal to those of a second or third class hotel in America .*
- Opinion: *As for myself , **I hate** Viareggio at all seasons*

They are often marked by linguistic expressions of subjectivity related to the narrator or to another character.

> *This is another " Cornice Drive , " and **far finer** ,// **I think** //, than that along the Riviera In point of comfort there **seems to us** little difference between first and second class;*

The presence of evalutative and emotive adjectives can help in identifying the EVALUATIVE nature of a clause.

> *the fare is **better** at the first , the company at the other ... .*

> *A **sad** little incident , under the pines ... .*

Exclamations and rhetorical questions can be annotated as evaluative.

> *It has not even a central piazza!*

> *Is it not a sign of empty-headedness ?*

Clauses introducing an EVALUATIVE clause and that express subjectivity (e.g. *I think*, *I believe*), must be annotated as NARRATIVE.

The attribute CONVERSATIONAL should be used to specify if a clause is the content of a direct or reported speech. Indeed this attribute has two values:
- DIRECT: indicates that the annotated clause is part of a direct speech.
  *"Rather an **extreme** measure ," // I suggested. // "It is . // Not that I **love** the saints . // But I **love** the police still less ."*
- REPORTED: indicates that the annotated clause is part of a reported speech.
  *He says // the trains are **so slow** and the waits **so long** // that we make almost as **good** time by driving , // and then it is **much more interesting** .*

In all the remaining cases, leave this attribute blank.

Please note that in the examples above, "*I suggested*" and "*He says*" are to be annotated as NARRATIVE.

Note that speculations on possible/probable outcomes of events, or causes/enablements of events, unless they explicitly contains an evaluative marker of subjectivity should be annotated as NARRATIVE:

- *I think// tomorrow it will rain* = NARRATIVE // NARRATIVE
- *Oshkosh Truck attributed the downturn in its earnings to higher start - up costs of its new chassis division , a softer motor - home market and higher administrative costs of compliance with government contractor regulations .* = NARRATIVE
- *I think // tomorrow will be a good day* = NARRATIVE // EVALUATIVE
- *Castro has said // that you can be a Communist // and still be a Christian .* = NARRATIVE // DESCRIPTIVE // DESCRIPTIVE
- *The main negative is // the risk that the Pope 's visit will persuade a great many more Cubans // to break loose of the Cuban government* = EVALUATIVE // EVALUATIVE // EVALUATIVE → Note that in this example the second and third clause are in the scope of an evaluative statement (The main negative is) which influence the class assignment of the following elements
- *It will give the rest of the world the view // that Cuba is like any other nation , // something the US has , of course , tried // to persuade the world // that it is not .* = all clauses are NARRATIVE

3.3 DESCRIPTIVE

Clauses in which the characteristics of an object, person or location are presented creating a mental picture in the reader's mind. These characteristics can be both visible and invisible, both tangible/intangible.

She is a spirited creature , but with a fine balance of common sense -
invisible/intangible
Over one of the doors is a Virgin and Child - visible/tangible

They often include an high quantity of adjectives and, especially in case of visible characteristics, also linguistic elements expressing spatial order (*left*, *top*, *in the background*)
The road winds above, beneath, and beside rugged cliffs of great height, always with the sea in full view .

Adjectives used in the DESCRIPTIVE text type are not evaluative (e.g. *good*) or emotive (e.g. *horrible*) adjectives: these are typically related to the EVALUATIVE text type. On the contrary, adjectives expressing the size, color, or shape of a person, a thing, an animal, or a place can be found in the DESCRIPTIVE type.
the sea was high

DESCRIPTIVE clauses often refer to qualities (e.g. color / look) and parts (for example when describing a building, its different rooms can be mentioned).
hers [apartment] is on the second floor , with a huge salon on one side of the hall , a salle à manger on the other , and the usual complement of bedrooms , kitchens , and the like .

Roles expressed in copula constructions with the verb "to be" can be annotated as DESCRIPTIVE:
Yvon Bussieres was the senior executive vice president.

DESCRIPTIVE clauses typically do not contain dynamism, are all expressed in the same reference time and contain states. In case of explicit temporal references (temporal expressions as well as temporal adverbs or other temporal signals) or verbs encoding a change of state such clauses become NARRATIVE, see the difference between the following two examples.
Change of state: *Yvon Bussieres was **named** senior executive vice president*
Temporal expression: ***Last year** Yvon Bussieres was senior executive vice president*
Temporal signal: *The sun was hot **while** arriving in Naples.* (a DESCRIPTIVE would be: *In the garden the sun was hot and roses were blooming everywhere).*

The attribute TYPE is used to distinguish between descriptions of: persons, locations and other.

- Locations are geopolitical entities, facilities (man made structures), other (referring to a physical location but different from the previous ones).
  Vesuvius, Capri, Ischia, and all the smaller islands of the bay were bathed in heliotrope light, a royal array of purple velvet.
- Persons are single individuals or group of humans.
  the second vetturino proved to be a better linguist than his predecessor
- The "Other" value is to be assigned to all the descriptions not covered by the previous options, e.g. an object, an activity, an organization:

object: *her black dress was coarse but tidy*
activity: *Being commander means // supervising the rest of the crew in training*
organization: *Telerate provides information about financial markets through an electronic network .*

As for the previous content type, also clauses of type DESCRIPTIVE has the CONVERSATIONAL attribute.

3.5 EXPOSITORY

Generalizations with respect to a class should be marked as Expository/Informative. These clauses include linguistic expressions that make statements about or refer to kinds giving general information about such kinds.

Roughly 100 children die in an average flu season .
Bricks are cold , stone is cold , tiles are cold

As for the previous content type, also EXPOSITORY has the CONVERSATIONAL attribute.

3.6 INSTRUCTIVE

Clauses containing procedural information for example the steps to be followed in a tour to reach a specific place.

At last you cross that big road // and strike the limestone rock .

As for the previous content type, also clauses of type INSTRUCTIVE has the CONVERSATIONAL attribute.

TWA eight hundred climb // maintain one five thousand
TWA eight hundred , if you hear center , ident

3.7 NONE

Clauses whose texts have mainly a structural purposes, e.g. headers and titles.
NAPLES , March 7th .

Often headers and titles contain the name of the location that is the main topic of the document and a date. Locations must be annotated with the LOCATION markable and dates with the TIMEX3 markable. TIMEX3 has 2 attributes: value and dct.

- Value: expresses the meaning of a temporal expression, i.e. it assigns to the temporal expression a normalized value corresponding to, for example, a calendar date. Its format follows the ISO 8601 standard: YYYY-MM-[WW]-DDThh:mm:ss, that is Year, Month, Week (optional), Day, Hour, Minute and Second. The value should be manually written in the text field. In case of unspecified components of a calendar date, use the placeholder character, X, for each unfilled position in the value of a component.

- DCT: has two options. If yes is selected this means that that TIMEX3 should be used as the document creation time and thus as an anchor for the other temporal expressions present in the document.

  CAPRI, March 13th .
  CAPRI = LOCATION
  March 13th = TIMEX3, value = XXXX-03-13, DCT = yes

3.8 OTHER

Clauses that cannot be labeled with the previous categories are to be marked with "Other". This markable has an attribute, named TYPE, to indicate the specific type of clause, if possible.

- literary_reference: such as citations of poems, transcriptions
  " ' My soul to-day Is far away , Sailing the Vesuvian Bay ; My winged boat , A bird afloat , Swims round the purple peaks remote . ' "
- phatic: phatic expressions (usually surrounded by commas "I mean", "you know")
- reader: references to the reader
  Your most interesting letter , Sir Philosopher , reached me at Gibraltar
  Please note that some authors use "you" to refer to an unspecified person: in this case clauses are not to be annotated as OTHER with TYPE=reader because the second-person pronoun is an impersonal/indefinite you.
- question: (rhetorical) questions not expressing an evaluation
  And how do we get on together?
- other_language: clauses in languages different from English
  Madame est servie.
- na: when all other options are not valid. E.g. exclamations (*Not very often indeed!)*,

As for the previous content type, also clauses of type OTHER has the CONVERSATIONAL attribute.

4. PRIORITY

In some cases a clause has a content that could be assigned to more than one text type: the mode with the highest priority in the following list should be assign that to the clause.

1. NARRATIVE
2. EVALUATIVE
3. DESCRIPTIVE
4. EXPOSITORY
5. INSTRUCTIVE
6. OTHER

The following example is a mix of EVALUATIVE and DESCRIPTIVE. On the basis of the previous priority list, the clauses have to be annotated as EVALUATIVE.

There is no commerce to liven them up // and make them smilingly polite ; //no historical tradition to give them self-respect ; //no agriculture worth mentioning //(the soil is too poor)//- in other words , no peasantry to replenish the gaps in city life //and infuse an element of decency and depth.

Next example is a mix of DESCRIPTIVE and OTHER, i.e. a literary reference, but the first option is to be preferred.

Many of them answer to Hawthorne 's description of the streets of Perugia // which are " like caverns //, being arched all over ..."

5. RULES OF THUMBS

- Parenthetical clauses are to be annotated on the basis of their functional and semantic content thus any content type label can be applied. The only exception is when such clauses are phatic thus they have to be annotated as OTHER.
- Punctuation must be included in the markable extent.
- Affective verbs (e.g. *hope*/*love*/*hate*) are typically in EVALUATIVE clauses.