

Attribution Relations Annotation Guidelines

VUA Perspectives Group

January 15, 2018

Contact person: Roser Morante (r.morantevallejo@vu.nl)

Contents

1	Attribution Relations: A Definition	3
2	Attribution Relation Markables Annotation	4
2.1	The <AttrSource> tag	5
2.1.1	The <AttrSource> Attributes	7
2.2	The <AttrCue> tag	7
2.3	The <AttrContent> tag	10
2.4	The Attribution Links	11
2.5	Additional AR Examples	11
2.5.1	Legal language	11
2.5.2	Call	11
2.5.3	Hear	11
2.5.4	Know	12
2.5.5	Prompt	12
2.5.6	Refuse	12
2.5.7	Say	13
2.5.8	Seek	13
2.5.9	Speak	13
2.5.10	Tell	13
2.5.11	To be concerned	13
2.5.12	View	14
2.5.13	Want	14
2.5.14	Passive construction	14
2.5.15	Imperative construction	14
2.5.16	Infinitive constructions with <i>to</i>	14
2.5.17	Present participle constructions	15
2.5.18	Bibliographic reference as SOURCE	15
2.5.19	Cases with no ARs	15

3	Annotation in eHOST: Instructions for Annotators	15
3.1	Configuring the Workspace	16
3.2	Introduce annotator name	18
3.3	Selecting a project	18
3.4	Annotating markables	19
3.5	Annotating attributes of markables	20
3.6	Discontinuous span for markables	20
3.7	Adjusting the span of a markable	21
3.8	Annotating relations	22
3.9	Saving your data	24
A	Appendix: Lists of potential noun cues (Pareti, 2015b)	25
B	Appendix: Lists of potential verb cues with number of occurrences as cue in the PARC corpus (Pareti, 2015b)	26

The Penn Discourse TreeBank (PDTB) (Prasad et al., 2006) and the Penn Attribution Relation Corpus (PARC) (Pareti, 2012) are the two reference corpora for Attribution Relations (ARs). In this document we present the annotation guidelines for ARs. These guidelines are based on the PARC v3.0 annotation guidelines (Pareti, 2015a).

1 Attribution Relations: A Definition

The working definition that Pareti (Pareti, 2015b) adopts is that of “a relation ascribing the ownership of an attitude towards some linguistic material, i.e. the text itself, a portion of it or its semantic content, to an entity”. An AR is established when someone (the author of a text, someone mentioned in the text) signals an ownership relation between a third party who expresses an attitude and some text. For example, the authors of sentence “Students like annotation tasks” establishes a relation between a source (students) and a content (annotation tasks) by means of a cue (like).

Following the PARC v3.0 annotation guidelines, the text towards which an attitude is expressed can be:

- An utterance or an assertion (i.e. either reported speech, or indirect reported speech, or mixed reported speech) as in “The student said *that he would annotate*”.
- A belief, as in “the teacher thinks *that the annotation task is easy*”.
- A knowledge, as in “the task leader knows *everything about attribution*”.
- An intention, as in “the student aims *at annotating three texts per day*”. It can be the case that either an agent influences someone to perform - or not - an action, that an agent commits to performing - or not - an action, or that an agent expresses a (mental) orientation towards some state of affairs.

An AR is typically expressed by three components:

- A SOURCE: the owner of the attitude. A SOURCE is typically an agent (animate or inanimate). It can be realized by Named Entities (PERSON, ORGANIZATION or COUNTRY), Proper Nouns, or Noun Phrases (NPs) (indefinite NPs, full definite NPs, or pronouns). In the following examples all sources are marked in bold; e.g.:
 1. **John** ordered the children to go away.
 2. **China** claimed the possession of the Senkaku islands.
 3. **People** blame the park for the measles outbreak.
 4. **He** said he would come.
- A CUE: a lexical item (a verb, a noun, a Prepositional Phrase (PP), an adjective or an adverb) which explicitly signals the ownership relationship between a SOURCE and a CONTENT. It can also signal the type of AR (e.g. assertion or belief). In the following examples all cues are marked in bold; e.g.:
 5. It is **said** that people are scared about vaccines.
 6. John **ordered** the children to go away.
 7. China **claimed** the possession of the Senkaku islands.
 8. People **blame** the park for the measles outbreak.

9. He **said** he would come.

- A **CONTENT**: a text portion which “is perceived as meant to be attributed to the **SOURCE**.” The **CONTENT** is the linguistic material which is attributed to the source. In the following examples all contents are marked in bold; e.g.:

10. It is said **that people are scared about vaccines**.

11. John ordered **the children to go away**.

12. China claimed **the possession of the Senkaku islands**.

13. People blame **the park for the measles outbreak**.

14. He said **he would come**.

Elements which may signal additional relevant information, such as the recipient of an order or of an assertion, information concurring to the identification of the provenance of a content, event specifications of time and place, or further specification of an attitude of the source, and similar are NOT annotated. We depart from the PARC v3.0 guidelines with respect to this points¹. This is why in the sentence (15), *patiently to the patient on Friday* does not have to be annotated, whereas the **SOURCE** (*the doctor*), **CUE** (*explained*), and **CONTENT** (*that she did not need to take more medication*) are enumerated.

15. **The doctor explained** patiently to the patient on Friday **that she did not need to take more medication**.

2 Attribution Relation Markables Annotation

An AR is expressed by at most 3 markables:

- **<AttrSource>**: this markable is used to annotate all mentions of **SOURCES** in an AR as defined above. It corresponds to the PARC v3.0 **<Source>** markable;
- **<AttrCue>**: this markable is used to annotate all mentions of **CUES** in an AR. It corresponds to the PARC v3.0 **<Cue>** markable;
- **<AttrContent>**: this markable is used to annotate all mentions of **CONTENTS** in an AR as defined above. It corresponds to the PARC v3.0 **<Content>** markable;

For an AR to occur at least the **CONTENT** of the AR has to be expressed in the text. The **CUE** and the **SOURCE** do not have to be expressed for an AR to occur. In the following example there is no AR because the **CONTENT** of the AR introduced by the **CUE** *believe* is not explicitly expressed.

16. “This community as a whole is not nearly as pristine and as efficient in its practices as you seem to believe”.

Three different AR links, are used to connect the three components of an AR:

¹In PARC these elements are annotated with the **<SUPPLEMENT>** tag

- `CONTENT_CUE_LINK`: links a `CONTENT` with a `CUE`.
- `CONTENT_SOURCE_LINK`: links a `CONTENT` with a `SOURCE`.
- `ONLY_CONTENT_LINK`: links a `CONTENT` with itself when no `CUE` and `CONTENT` are available.

Below we explain the markables in more detail.

2.1 The `<AttrSource>` tag

The `<AttrSource>` tag is the markable used to annotate sources in an AR. The source of an AR is any entity (animate or inanimate) which has the role of an agent and owns a specific attitude towards a certain linguistic material expressed by the content.

The annotation of `<AttrSource>` follows the instructions reported in the PARC v3.0 guidelines concerning the span extent. The span of the `<AttrSource>` tag corresponds to the maximal Noun Phrase realizing the source in the text. Appositions and relative clauses which contribute to the characterization of the source must be included in the tag span (including the punctuation marks like commas which can visually characterize these text spans.) e.g.:

17. He said he would come.
`<AttrSource>He</AttrSource>` said he would come.
18. Mr Potts, the owner of Pottery Inc., refused to pay his debts.
`<AttrSource>Mr Potts, the owner of Pottery Inc.,</AttrSource>` refused to pay his debts
19. "Unbelievable!" said John Harris, a university official who is working on gravitational waves.
 "Unbelievable!" said `<AttrSource>John Harris, a university official who is working on gravitational waves</AttrSource>`.

In case the source element is realized by an adjective,² the maximal NP must be annotated as well. Exceptions to this rule apply to possessive adjectives: in case a possessive adjective modifies a potential noun which can function as a cue, then only the possessive adjective must be marked with the `<AttrSource>` tag, as in Example (21). If a possessive adjective modifies a noun which cannot function as a cue, then the full NP is marked, as in Example (22).

20. `<AttrSource>The presidential report</AttrSource>` shows that the amount of digitized data has increased a lot.
21. `<AttrSource>His</AttrSource>` advice: "Run!"
22. According to `<AttrSource>his report</AttrSource>` the figures are expected to increase.

In case the source element is realized by a relative pronoun and the AR is realized in a relative clause, then only the relative pronoun must be annotated as source. Anaphoric relations are out of the scope of this task; e.g.:

²i.e. a premodifier of a NP with a (coreferential or associative) anaphoric link to a source previously expressed in the text or which makes the source implicit in the text available among the entities in the Universe of the Discourse.

23. The bank, <AttrSource>which</AttrSource> said it could not be able to repay all of its debt, has lost 10% of its value on Wall Street.

Coordinated sources Multiple sources in a coordination relation must be annotated as separate mentions and will give rise to separate ARs. In the following sentence there are two cases of attribution, one for the SOURCE *Germany* and another for the SOURCE *France*, even if the CONTENT is the same in both ARs.

24. Both <AttrSource>France</AttrSource> and Germany would <AttrCue>like</AttrCue>
<AttrContent>to boost the EU's foreign policy</AttrContent>
25. Both France and <AttrSource>Germany</AttrSource> would <AttrCue>like</AttrCue>
<AttrContent>to boost the EU's foreign policy</AttrContent>

Underspecified sources are those in which the source of an AR is explicitly realized by an NP or another lexical item that express a referent (e.g. “some people”; “rumors” etc) which cannot be associated with any external referent in the world. Underspecified sources will be annotated:

26. <AttrSource>Rumors</AttrSource> say that ...
27. <AttrSource>Some people</AttrSource> report ...
28. <AttrSource>Officials from the Pentagon</AttrSource> deny ...

Implicit sources are sources which have a linguistic realization in the text (e.g. a pronoun), which cannot be associated with any specific referents or be resolved with any previously mentioned entities. In other cases, the use of a particular syntactic construction (e.g. passive) and a cue is used to signal the presence of these implicit sources. In case there is no lexical item realized in the text, then the SOURCE will not be annotated, but the other components yes, as in the following example where the passive construction allows for an implicit source.

29. <AttrContent>Prices</AttrContent> are <AttrCue>expected</AttrCue>
<AttrContent>to rise</AttrContent>.

Sources are intentional entities or have been intentionally created The source (animate or inanimate) MUST express an intention or a volition. Special cases of sources such as *news reports*, *documents*, *reports* and similar entities are annotated because they have been intentionally created. Attribution relations are NOT evidential relations. Bare evidences (e.g. figures from a statistics) should not be considered or annotated as sources: there is no intention or volition in “making them available, the intention is normally to use them to support some line of reasoning. Compare the two examples below (taken from PARC v3.0 guidelines). In Example (30) *report* is considered to be a SOURCE, whereas in Example (31) *figures* are not considered to be a SOURCE.

30. <AttrSource>The report</AttrSource> shows that deaths on urban interstate highways rose 7% between 1986 and ...
31. The figures from the past years show that deaths on urban interstate highways rose 7% between 1986 and ...

2.1.1 The <AttrSource> Attributes

The <AttrSource> tag has one attribute: `Source_type`, which is a non-optional attribute. Every time a source is annotated, the type of the source must be specified. By type of the source, we are not referring to the entity type like in Named Entities tasks (PERSON, ORGANIZATION, etc.) but rather to the nature of the source regarding its semantic transparency in the document. The values of the attributes are the following:

- **Explicit:** this value is used every time a third party source is introduced in the text. An explicit source is a source whose identity can be easily retrieved and identified as unique. Explicit sources do not necessarily correspond to entities linked to structured knowledge repositories like DPpedia or Wikipedia. An explicit source corresponds to a lexical item or an NP whose referent in any possible world can be identified. In other terms, the source can be identified as a `Character`.

32. **The note** states that ...

33. **The president of the U.S.A** claimed ...

34. **He**³ wanted to go ...

35. **A spokesperson of the White House** ...

36. **Germany** recently announced plans to fine parents who failed to get medical advice about immunising their child.

- **Generic:** this value must be assigned to underspecified and generic sources. A source is generic when it is not possible to identify the specific entity to which it refers. For example, if a source is “officials from the Pentagon”, it is not possible to identify who are the specific officials that the expression refers to. The same applies to the following examples:

37. **Rumors** say that ...

38. **Some people** report ...

39. **Officials from the Pentagon** deny ...

40. **37 states** approved the law

- **Author:** this value is to be used anytime that the author of the document is the owner of the attitude being expressed by the AR.

2.2 The <AttrCue> tag

The <AttrCue> tag is used to annotate lexical items which explicitly introduce an AR. In terms of parts-of-speech, cues can be realized by:

- Verbs: the following cases illustrates when a verb functions/may function as a cue of an AR (Pareti, 2015a):
 - Attributional verbs (e.g.: verbs of reporting, knowledge, beliefs, verbs of influence, verbs of commitment, verbs of orientation);

³The pronoun can be resolved to a unique explicit entity.

- Attributional verbs entailing manner (e.g.: *quip, grouse, brag* etc.)
- Manner verbs entailing a reporting verb (e.g.: *fume*)
- Verbs entailing the manner of an implicit co-temporal reporting verb (e.g.: *smile, sigh*, etc.)
- Nouns:
 - 41. The investors had *expectations* that the market would recover.
- Adverbs: *reportedly, allegedly*.
 - 42. The two companies *reportedly* plan launch a new product this year.
- Adjectives:
 - 43. I am *sure* they will cancel the flight.
- Graphic markers such as :
 - 44. Trump: “I will cancel Obamacare”.
- Prepositional groups/Prepositions: *according to; in the eyes of, ...*)
 - 45. *In the eyes* of the Catalan government a referendum was necessary.

For examples of potential AR cues see the lists of Attribution Verbs and Attribution Nouns from Pareti (2015a), included here in Appendix A.

Cues are not always expressed in an AR, they can be left unrealized: in these cases we will deal with ARs composed by only CONTENT or a pair of SOURCE and CONTENT.

Annotating cues A cue normally realizes an event. In compliance with other annotation layers provided in this scheme, the annotation of cues is performed following the event annotation guidelines. In particular:

- In case the cue is realized by a verb, noun or adjective, follow the minimal chunk approach: annotate the phrase head, excluding all pre- and post-modifiers, articles, auxiliaries (including modals), negative particles, adverbials and similar; e.g.:
 - 32. The report has <AttrCue>shown</AttrCue> that deaths on urban interstate highways rose 7% between 1986 and ...
 - 33. Our <AttrCue>hope</AttrCue> is that the Senator will leave office.
 - 34. He did not <AttrCue>say</AttrCue> he will ...
 - 35. I am <AttrCue>hopeful</AttrCue> that I will get a promotion.
 - 36. The Senate is <AttrCue>expected</AttrCue> to reach an agreement today
- In case the cue is realized by a prepositional group or a preposition, always annotate the full prepositional group; e.g.:
 - 37. <AttrCue>According to</AttrCue> Mary John must leave now.

- Punctuation markers such as semicolons, commas, and similar MUST be annotated only when NO OTHER LEXICAL CUE is available. Quotation marks are NOT part of the cue span, but they are part of the content, as proposed in the PARC v3.0 guidelines; e.g.:

38. His <AttrCue>advice</AttrCue>: “Run!”

39. The MP was delusional <AttrCue>:</AttrCue> “I will rule the world!”

Coordinated sources Following the PARC v3.0 guidelines, **one cue can be associated to only one AR**. An exception to this are cases with coordinated sources:

40. Both <AttrSource>France</AttrSource> and Germany would <AttrCue>like</AttrCue>
<AttrContent>to boost the EUs foreign policy</AttrContent>

41. Both France and <AttrSource>Germany</AttrSource> would <AttrCue>like</AttrCue>
<AttrContent>to boost the EUs foreign policy</AttrContent>

Coordinated cues Multiple cues in coordination relation must be annotated as separate mentions and will give rise to separate ARs although the source and content may be the same.

42. <AttrSource>Analysts</AttrSource> <AttrCue>know</AttrCue> and believe
<AttrContent>that the market is at a turning point</AttrContent>.

43. <AttrSource>Analysts</AttrSource> know and <AttrCue>believe</AttrCue>
<AttrContent>that the market is at a turning point</AttrContent>.

Cases in which the cue is not present In some cases the cue could be implicitly stated or not lexically realized. Textual genres may influence such cases. In news articles, a general case of missing cue in an AR is represented by isolated sentences which report a direct quotation. To clarify, consider the annotation of the following two sentences from PARC v3.0:

44. As the crowd [...] shouted “ANC, ANC” the old man shot his fists into the air.
“I am inspired by the mood of the people”. (PARC v3.0 annotation - wsj_2454.xml)

In this short excerpt, there are two attribution relations, which must be annotated as follows (notice the rule: one cue, one AR):

- AR1: <AttrSource> - the crowd; <AttrCue> - shouted; <AttrContent> - “ANC, ANC”
- AR2: <AttrSource> - the old man; <AttrContent> - “I am inspired by the mood of the people”

2.3 The <AttrContent> tag

The <AttrContent> tag is used to annotate all mentions of content. Following the PARC v3.0 guidelines, as a general principle the annotation of the content should be limited to that portion of the text that is perceived as meant to be attributed to a source.

General rules that apply for the annotation of CONTENT are the following ones:

- Quotation marks, when present, must be included in the tag span;

45. <AttrSource>The old man</AttrSource> shot his fists into the air.
<AttrContent>“I am inspired by the mood of the people”</AttrContent>.

- Content span may be discontinuous and span over sentence boundaries;

46. <AttrContent>“Having the dividend increases is a supportive element in the market outlook, but I don’t think it’s a main consideration”</AttrContent>,
<AttrSource>he</AttrSource> <AttrCue>says</AttrCue>.
<AttrContent>Within slower economic growth and flat corporate earnings likely next year, “I wouldn’t look for the market to have much upside from current levels</AttrContent>.

- The complementizer “that” must always be included in the content span;

47. As the crowd [...] shouted “ANC, ANC” the old man shot his fists into the air.
“I am inspired by the mood of the people”.

- Punctuation marks at the end of a content span must be included only if part of the content, otherwise it can be excluded; e.g.:

48. <AttrContent>The Senate</AttrContent> is <AttrCue>expected</AttrCue>
<AttrContent>to reach an agreement today</AttrContent>. [punctuation mark excluded]

49. <AttrSource>He</AttrSource> <AttrCue>said</AttrCue>:
<AttrContent>“I will rule the world!”</AttrContent> [quotation marks and punctuation mark included]

50. <AttrSource>I</AttrSource> am <AttrCue>wondering</AttrCue>
<AttrContent>how is that possible?</AttrContent> [punctuation mark included]

Content can be realized also by means of anaphoric relations with previously stated/asserted/believed etc. clauses or sentences in the text. In this case, only the anaphoric element must be annotated as content. The connection with the antecedent will be performed as part of an anaphora resolution task - out of the scope of this task. In Example (51), *That connection* is the anaphoric element.

51. But in the developed world it’s an artifact of the anti-vaccination movement, which has associated the vaccine with autism. <AttrContent>That connection</AttrContent>, promoted by the discredited British physician Andrew Wakefield and the starlet Jenny McCarthy, has been thoroughly <AttrCue>debunked</AttrCue>.

Content does not include the addressee When a verb requires a complement with an addressee, typical of communication verbs, the addressee should not be included in the content. In the example below *me* is excluded from the content.

52. The paediatrician <AttrCue>reassured</AttrCue> me <AttrContent>that they would be fine</AttrContent>.

2.4 The Attribution Links

The attribution links aim at creating an AR between the three markables, CONTENT, SOURCE, and CUE. CONTENT is the only compulsory element that has to be lexically realized for an AR to exist.

- The <Content_Source_link> is a link tag which connects <AttrContent> and <AttrSource> .
- The <Content_Cue_link> is a link tag which connects <AttrContent> and <AttrCue> .
- The <Only_Content_link> is a link tag which connects a <AttrContent> with itself when the SOURCE and CUE are not explicit.

The directionality of the links is from <AttrContent> to the other markables.

2.5 Additional AR Examples

In this section we report additional examples of annotation of attribution relations. To facilitate the reading of the examples we will mark in separate lines all markables, no link annotation will be provided.

2.5.1 Legal language

53. The court found him guilty.
SOURCE: The Court
CUE: found
CONTENT: him guilty

2.5.2 Call

54. “They are even called “immunizations.”” CUE: called
CONTENT: “immunizations”

2.5.3 Hear

55. John heard from Mary that he won the elections.
SOURCE: Mary

CONTENT: that he won the elections

56. John heard that he won the elections
No AR because nothing is being attributed to a source.

2.5.4 Know

57. I know that the outbreak started in Disneyland
SOURCE: I
CUE: know
CONTENT: that the outbreak started in Disneyland

58. I know a lot about vaccinations
No AR because nothing is being attributed to a source.

2.5.5 Prompt

59. Conspiracy theories about the health risks of certain vaccinations have spread on the internet, prompting some parents to shun immunisation.
No AR because *prompt* is being used with the meaning 'cause or bring about'.

2.5.6 Refuse

Refuse is annotated as an attribution cue when it is used to express a determination not to do something, as in the following example:

60. Mr. Smith said "I refuse to get out there and tell a false story".
SOURCE: I
CUE: refuse
CONTENT: to get out there and tell a false story
61. He refused to reveal the whereabouts of his daughter.
SOURCE: He
CUE: refused
CONTENT: to reveal the whereabouts of his daughter

Refuse is NOT annotated as an attribution when it means 'to decline to accept, to decline to give' as in the following examples:

62. The Big Board has refused to be drawn into a public debate.
63. Pittston refused to sign the pact.

2.5.7 Say

64. He said three words.

No AR because *say* is used with the meaning 'pronounce' and no content is specified.

65. "I am sorry". He said these three words.

SOURCE AR1: He

CUE AR1: said

CONTENT AR1: these three words

SOURCE AR2: He

CONTENT AR2: "I am sorry"

2.5.8 Seek

66. He seeks fame and fortune.

SOURCE: He

CUE: seeks

CONTENT: fame and fortune

67. She is seeking for clues to solve the case.

No AR.

2.5.9 Speak

68. He spoke the truth.

SOURCE: He

CUE: spoke

CONTENT: the truth

2.5.10 Tell

69. She told John to follow her.

SOURCE: She

CUE: told

CONTENT: John to follow her

2.5.11 To be concerned

70. Many people are concerned about the effects of climate change.

SOURCE: Many people

CUE: concerned

CONTENT: about the effects of climate change

2.5.12 View

71. My view today is that there likely have been some cases of measles caused by vaccines
SOURCE: my
CUE: view
CONTENT: that there likely have been some cases of measles caused by vaccines

2.5.13 Want

72. Marc wants to go home.
SOURCE: Marc
CUE: wants
CONTENT: to go home
73. He wants an ice-cream.
No AR because the object of *want* is a material object.

2.5.14 Passive construction

74. Asylum is claimed not just by people fleeing war, but also by people whose rights are being violated.
- SOURCE AR1: people fleeing war
CUE AR1: claimed
CONTENT AR1: asylum
- SOURCE AR1: people whose rights are being violated
CUE AR1: claimed
CONTENT AR1: asylum

2.5.15 Imperative construction

75. Ask yourself who the professional societies might be.
No AR because no attitude is being assigned to a SOURCE due to the imperative tense.

2.5.16 Infinitive constructions with *to*

76. It is wrong to assume they were unvaccinated.
CUE: assume
CONTENT: they were unvaccinated

77. "It's no use blaming unvaccinated persons for vaccine failures"
CUE: blaming
CONTENT: unvaccinated persons for vaccine failures

2.5.17 Present participle constructions

78. It's no use blaming unvaccinated persons for vaccine failures.
CUE: blaming
CONTENT: unvaccinated persons for vaccine failures

2.5.18 Bibliographic reference as SOURCE

79. "Vaccination and the use of antiviral drugs are two of the most important response measures for reducing morbidity and mortality during a pandemic." (WHO 2005d).

CONTENT: Vaccination and the use of antiviral drugs are two of the most important response measures for reducing morbidity and mortality during a pandemic
SOURCE: WHO 2005d (Notice: the brackets are not included)
CUE: ""

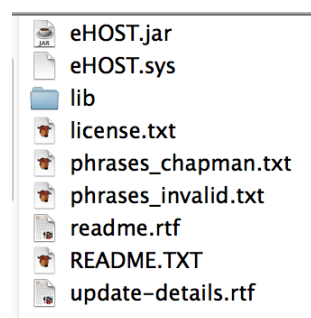
2.5.19 Cases with no ARs

80. No evidence suggests that the medicines are effective.

3 Annotation in eHOST: Instructions for Annotators

The annotator will first read the entire text of a document. After that she will proceed to annotate every AR that occurs in the text. She will annotate all relevant markables and relations of the AR at once.

For this annotation task we use the tool eHOST⁴. The documentation can be found at <http://ehostdoc.com/>. In order to install the tool, download the file ehost-bp-1.zip at <https://code.google.com/archive/p/ehost/downloads>, and unzip it in your preferred directory. As a result you will have a directory called ehost-bp-1 which looks like this:

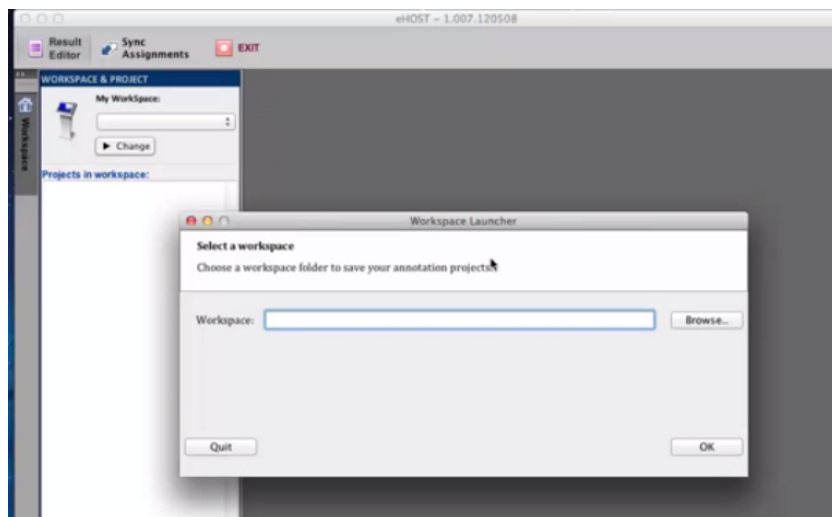


⁴<http://blulab.chpc.utah.edu/content/ehost-extensible-human-oracle-suite-tools>

You should also have a Dropbox account, which can be made at <https://www.dropbox.com>.

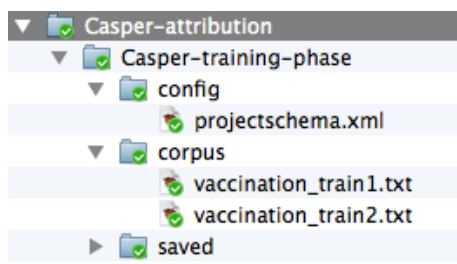
3.1 Configuring the Workspace

In order to launch the tool you will click in the file eHOST.jar. The first time you open eHost you will be asked to choose a workspace. A workspace is a folder with ongoing projects.



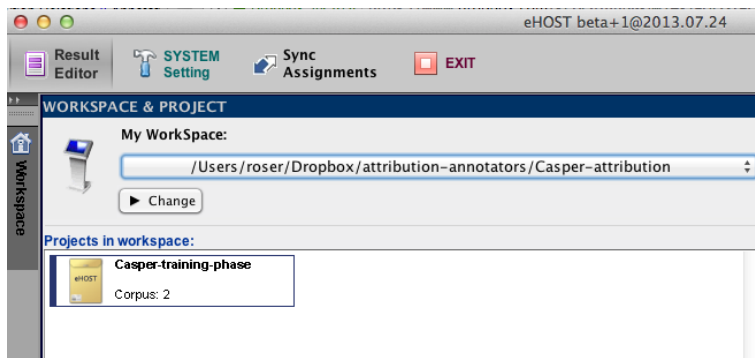
Watch now the tutorial video *Launching a workspace* to be found at <https://www.dropbox.com/s/1v3rtu8kiw14k/eHOST-Launch.mp4?dl=0>.

Configure your workspace to be the Dropbox folder that you will have been assigned by the task leader. The advantage of using Dropbox is that then your files will be stored in a safe place and the task leader can access them without you having to send them or upload them. Your workspace folder will have different projects. The first one will be one called “yourname-training-phase”. The folder should look like this:



The “config” subfolder contains the specification of the task, the “corpus” folder contains the files to be annotated, and the “saved” folder contains the saved annotations. When you start the task the “saved” folder is empty.

The selected workspace will appear in the tool interface. It will look like this:

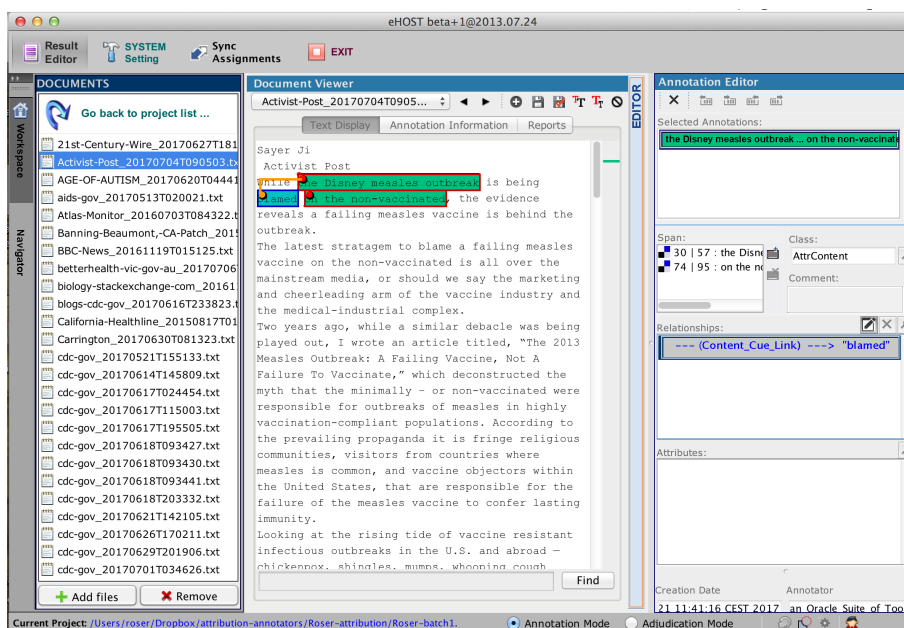


Once you have configured your workspace folder, you can start annotating. Before that, read Chapters 1 and 2 from the eHOST official Documentation Wiki (<http://ehostdoc.com/html/start.html>) and watch the following tutorial videos from the eHost website:

- *Annotating entities from a corpus* to be found at <https://www.dropbox.com/s/pja7jryxybfs9zg/eHOST-Relations.mp4?dl=0>
- *Annotating relationships between entities* to be found at <https://www.dropbox.com/s/pja7jryxybfs9zg/eHOST-Relations.mp4?dl=0>
- *Saving your annotated corpus and workspace* to be found at <https://www.dropbox.com/s/tdzgrtw8uf18o3d/eHOST-Save.mp4?dl=0\%20target=>.

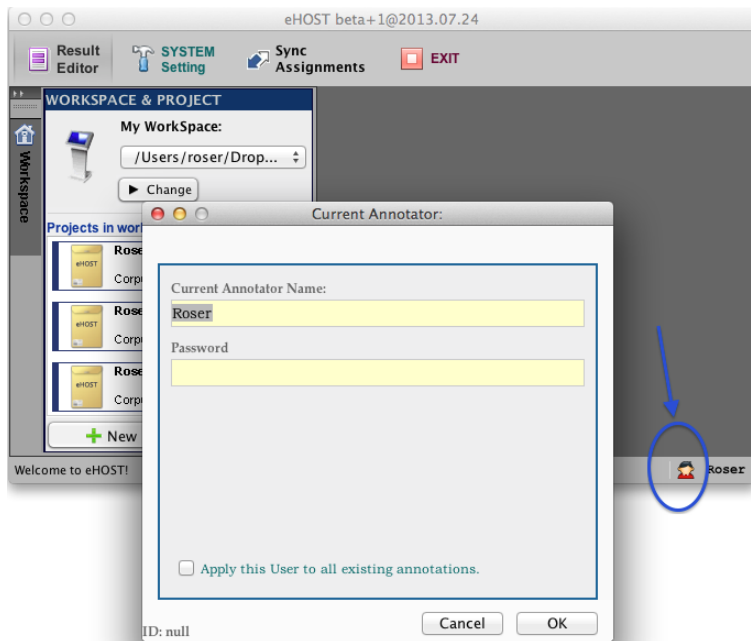
Additionally, you can watch the video *Demonstration of Annotation Tools* at <https://www.youtube.com/watch?v=T5kpPxxgw9tw> from minute 35:31 to minute 58:00.

The eHost Workspace interface is divided into three panels: Documents, which contains the list of files per project; Document Viewer, which displays the text to be annotated, and Annotation Editor, where several actions can be performed for a specific markable.



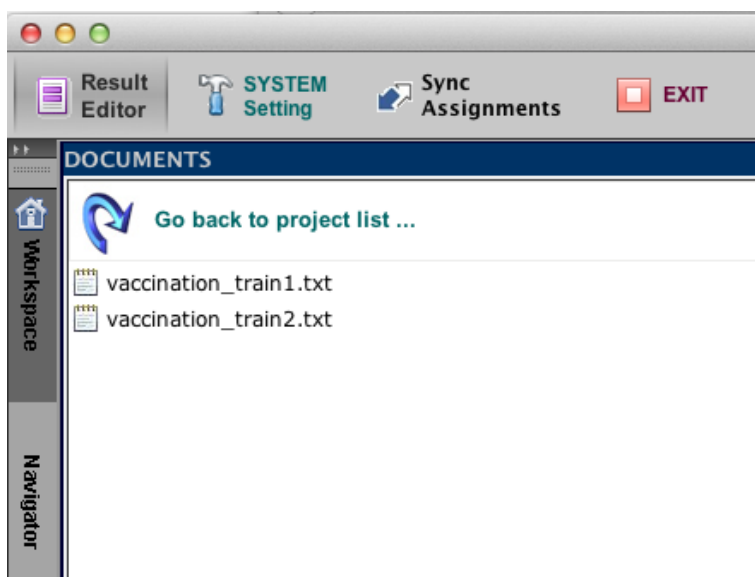
3.2 Introduce annotator name

Before you start annotating you should introduce your annotator name. You can do this by right-clicking at the person icon that appears in the right bottom corner of the interface. A pop-up window will appear where you can introduce the “Current Annotator Name”. Doing this is very important in order to be able to compare annotations by different annotators and calculate inter-annotator agreement.



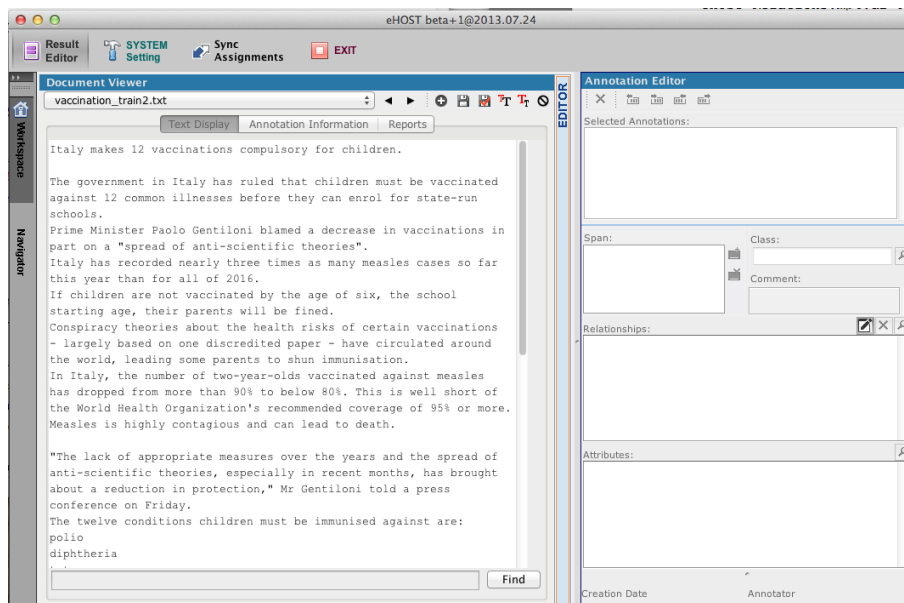
3.3 Selecting a project

In order to start the annotation task you need to select a project from the “Workspace and project” panel (see figure above). Double click on the project name. Let’s imagine that you select the project “Casper-training-phase” shown in a previous figure. You will then see a list of files in the “Documents” panel:



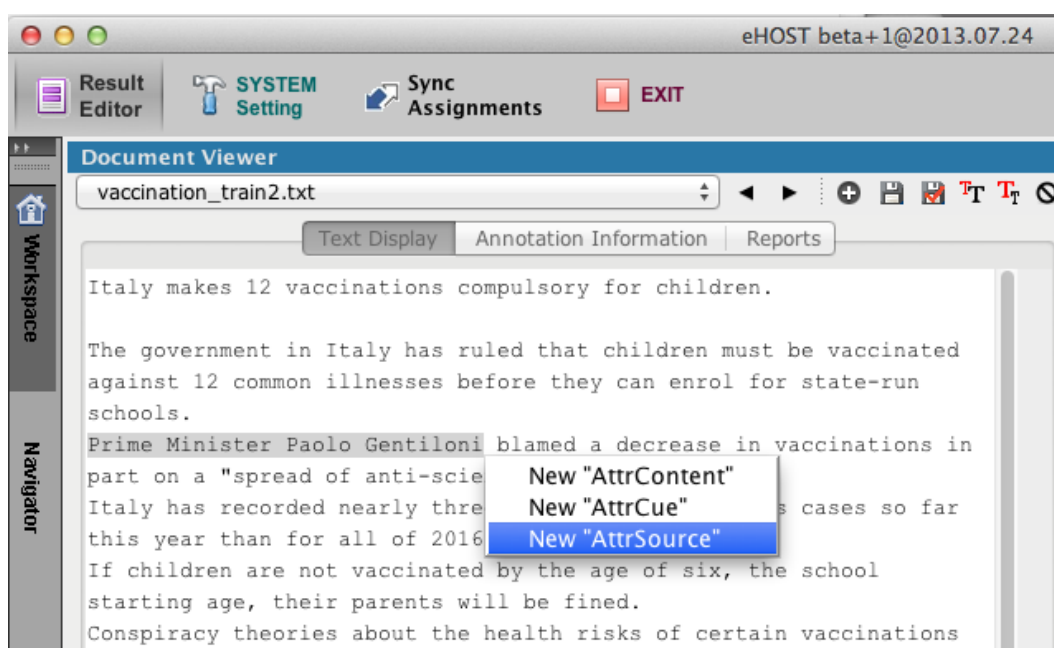
Double click in one of the files to start the annotation task. It will appear in the “Document viewer”

panel. First, read the full text once. Then proceed to annotate each AR.



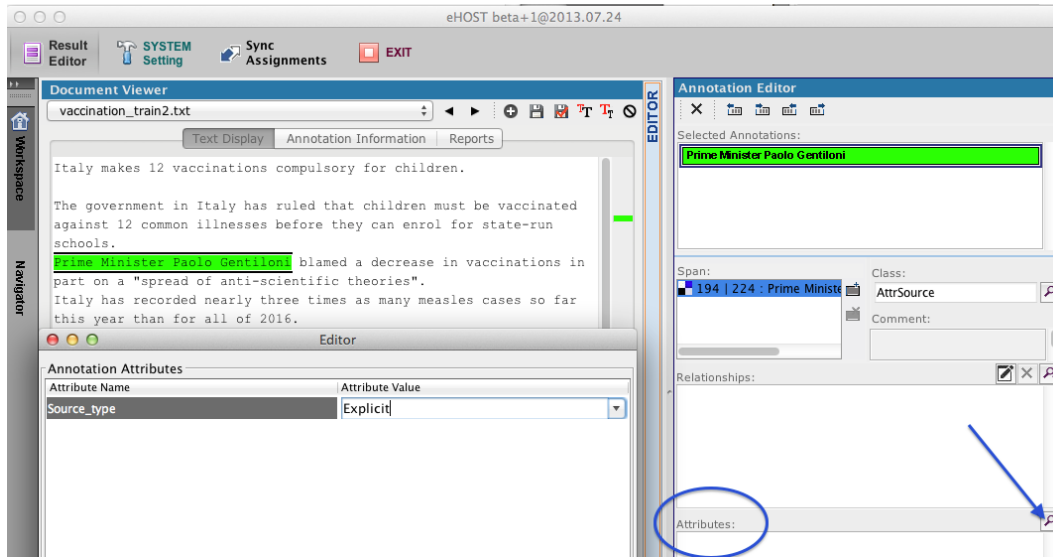
3.4 Annotating markables

For every AR you need to annotate three markables: `AttrSource`, `AttrCue` and `AttrContent`. As it has been indicated in the guidelines above, `SOURCE` and `CUE` might not be explicit in the text, whereas `CONTENT` has to be explicit for an AR to exist. Else, there is no AR. As you have seen in the video tutorials, in order to annotate you need to be in the “Text Display” tab of the Document Viewer panel. Select the text that you want to annotate with a markable. A pop-up window will appear where you can select the markable that you want to assign.



3.5 Annotating attributes of markables

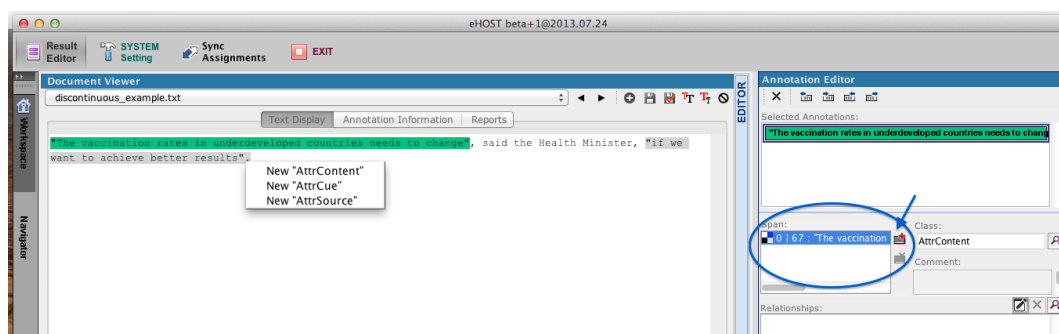
Additionally, for the `AttrSource` markable you have to assign a value of the attribute `Source_type`. You need to click on the markable and position the mouse on the Attributes box of the Annotation Editor panel. Click on the small search icon on the right and the Annotation Attributes window will pop up, where you can select an Attribute Name (`Source_type` in this case) and its value from a drop down menu. Choose one of the following values: `Explicit`, `Author`, `Generic`.



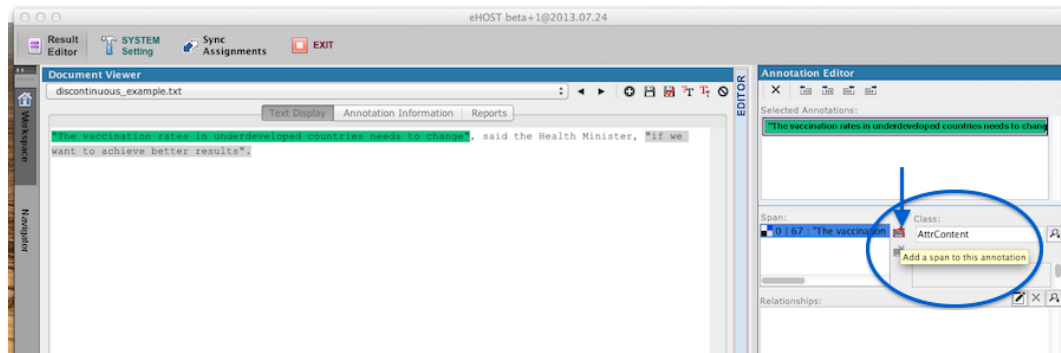
3.6 Discontinuous span for markables

It can be the case that a markable is discontinuous. In order to mark discontinuous spans of text you need to do the following:

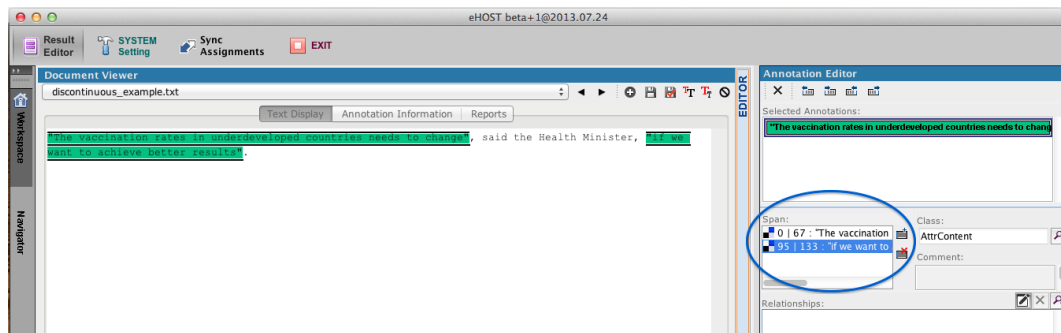
- Select the first part of the span and assign to it a class. The annotated span will appear in the Annotation Editor panel to the right of the interface.
- Then select another part of the markable span.



The menu to select a class will be displayed, but do not select a class. Go to the Annotation Editor panel, to the Span box where you will see the span of the first part of the markable that you annotated in the previous step.



To the right you will see an icon that is a folder with a + sign. Click on the icon and the span of the selected text will be added to the span of the previous step.

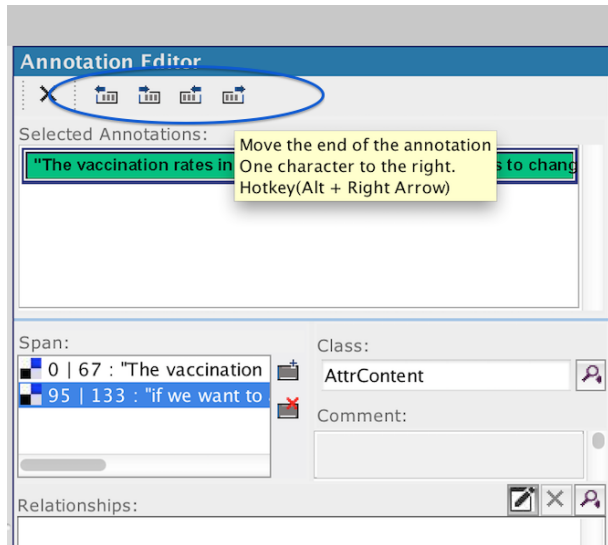


- Repeat the previous step as many times as necessary.

3.7 Adjusting the span of a markable

It might be the case that when you select text, the span of the markable is not correct. For example, you might want to exclude the final dot at the end of a sentence. The tool will first expand the selection to the punctuation mark attached to the word, but you can modify this by using the options in the Annotation Editor (the rectangles with arrows that appear below “Annotation Editor”).

Click on the markable whose span you want to change. Its span will appear in the Span box of the Annotation Editor. Click on the span that you want to change and press one of the 4 buttons for span modification, depending on how do you want to modify it. You can move the beginning and the end of the span one character to the right or to the left.



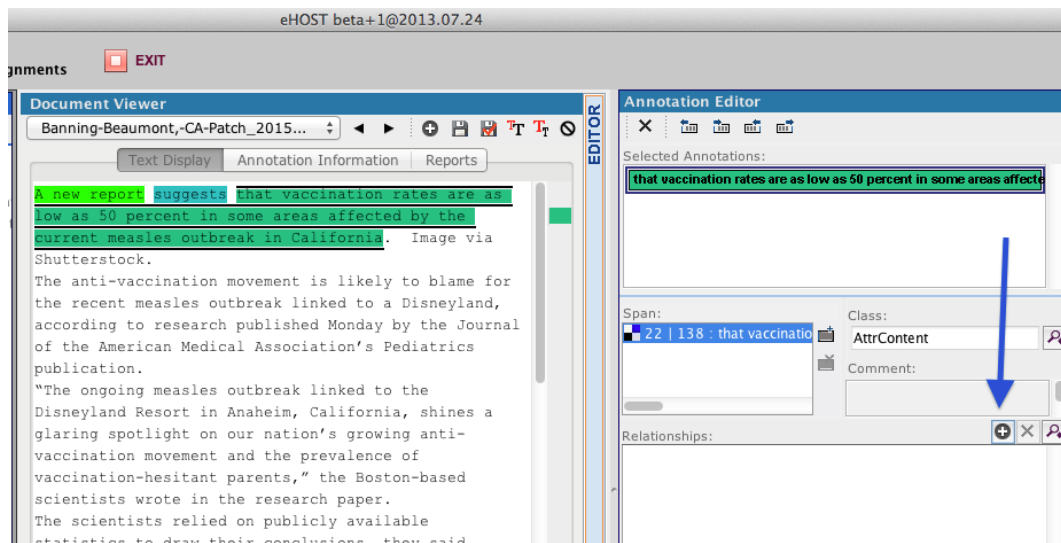
3.8 Annotating relations

Once the components of an AR have been annotated, it is necessary to annotate the relation between them. The tutorial video *Annotating relationships between entities* explains how to do this. Several relations are defined as markables for this annotation task:

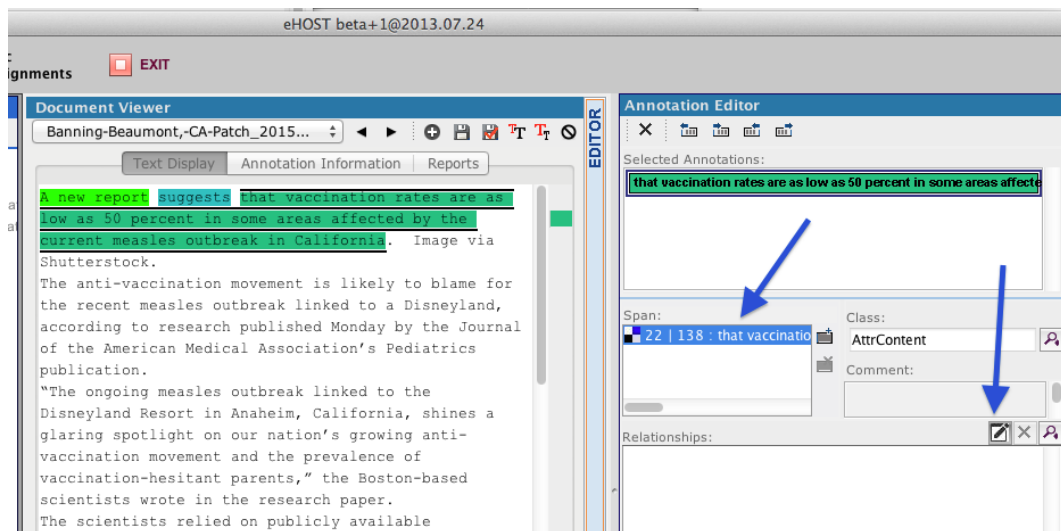
- The `<Content_Source_link>` is a relation markable which connects `<AttrContent>` and `<AttrSource>`.
- The `<Content_Cue_link>` is a relation markable which connects `<AttrContent>` and `<AttrCue>`.
- The `<Only_Content_link>` is a relation markable which connects a `<AttrContent>` with itself. This should be used only when both the SOURCE and CUE are not explicit.

The directionality of the relations is from `<AttrContent>` to the other markables. These are the steps:

- Select the `<AttrContent>` markable in Document Viewer.
- Click on the "+" symbol in the Relationships box of Annotation Editor.

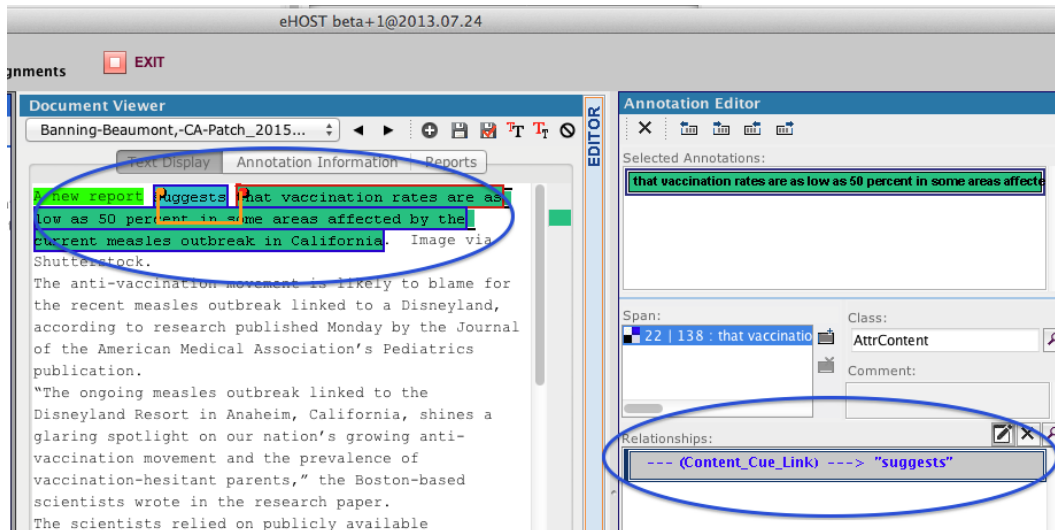


Now the Relationships box is in Edit mode:



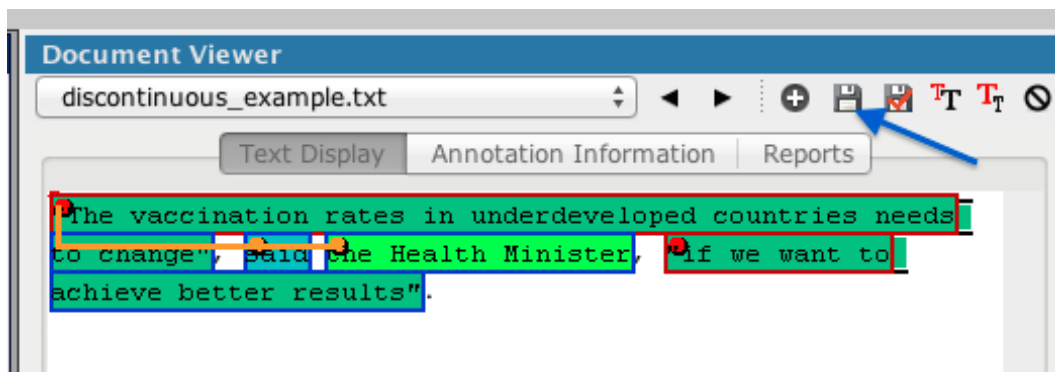
- In Document Viewer, right click on the other markable that participates in the relation.

The relation will appear in the Relationships box of Annotation Editor.



3.9 Saving your data

It is very important that you save your annotations frequently. eHOST does **NOT** save them for you. You can save them by clicking on the save button of the Document Viewer panel. The annotations will be saved in your Dropbox folder (assuming that you configured it to be your workspace). Thus, you do not need to send them to the task leader.



A Appendix: Lists of potential noun cues (Pareti, 2015b)

accord according accusation acknowledgement ad admission advice agreement allegation amendment announcement answer anticipation argument article assertion assumption assurance belief bet bill book call challenge charge chart citation claim command comment commercial complaint concern concession conclusion condition confession confidence confirmation consideration contention conviction counter-claim critic criticism cry data decision declaration deduction defence definition deliberation demand denial depiction description dictate disappointment disapproval disclosure discovery dispute document doubt effort elaboration encouragement eruption estimate eulogy evidence exclamation expectation explanation expression fear feeling figure file filing find finding forecast formulation guess highlight hint hope idea illustration implication imposition indication information insinuation inspiration instruction intention interjection interpretation issue joke knowledge lament laugh law lawsuit lecture legislation lesson letter list menace mention message mind moan need news note notice notification oath objection observation offer opinion order pact paper permission plan pledge point policy poll praise prediction press proclamation project promise proposal protest prove provision question quotation realization reason recognition recollection recommendation recount reflection reform refusal rejection remark repetition reply report reproach request requirement research resentment response revelation rule rumor saying scream shout sigh sign signal snort specification speculation spell statement statistic story strategy study suggestion suit support supposition survey suspicion talk temptation testimony theory thought threat understandment urge view voice want warning wisdom worry yell

B Appendix: Lists of potential verb cues with number of occurrences as cue in the PARC corpus (Pareti, 2015b)

absolve 1 accept 10 acclaim 1 account 1 accuse 50 acknowledge 46 add 372 address 3 admit 44 advise 23 advocate 2 affirm 2 agree 233 aim 2 allege 41 allow 28 allude 2 announce 186 answer 4 anticipate 21 appeal 2 applaud 3 appreciate 3 approach 2 approve 27 argue 121 ask 127 aspire 2 assail 2 assert 28 assess 1 assume 32 assure 6 attack 1 attempt 3 attest 3 attribute 46 authorize 8 avoid 1 await 3 back 2 ban 2 bar 2 battle 1 beam 1 become 1 beg 1 begin 1 believe 267 bemoan 1 bet 10 bid 1 bill 1 blame 21 block 3 boast 5 brag 2 brim 1 burble 1 buttress 1 buy 1 calculate 3 call 122 capture 1 care 1 caricature 1 castigate 1 caution 24 celebrate 2 challenge 3 characterize 4 charge 41 chastise 2 chide 1 choose 2 chuckle 1 cite 95 claim 84 clarify 2 clear 2 comment 26 commission 1 commit 2 communicate 1 compare 3 complain 55 comply 3 concede 45 concentrate 1 concern 20 conclude 39 concur 1 condemn 3 confess 6 confide 2 confirm 34 consent 10 consider 131 conspire 1 construe 1 contain 1 contemplate 2 contend 79 contest 1 continue 9 contribute 17 convey 1 convict 9 convince 18 couch 1 counsel 1 count 1 counter 4 credit 4 criticize 14 croon 1 crow 1 cry 1 dare 1 decide 66 declare 58 decline 48 decry 3 deem 4 defend 8 define 8 deflect 1 delight 1 deliver 1 demand 24 demonstrate 1 demur 1 denounce 3 deny 51 depict 1 deride 2 describe 34 desire 1 detail 1 detect 1 determine 12 develop 1 diagnose 1 dictate 3 direct 1 disagree 4 disappoint 3 disapprove 1 discern 2 discipline 1 disclaim 1 disclose 50 discourage 3 discover 14 discuss 20 disincline 1 dislike 1 dismiss 5 dispute 2 disturb 1 divide 6 doubt 16 downgrade 1 downplay 1 draw 1 dream 1 dub 3 echo 2 elaborate 2 embrace 1 emerge 1 emphasize 15 empower 2 enable 1 encourage 12 endorse 2 ensure 2 entice 3 envisage 1 envision 1 equate 3 erupt 1 establish 1 estimate 130 evaluate 1 evince 1 examine 1 exclaim 2 exclude 1 exhort 1 exonerate 1 expect 671 explain 47 explore 2 expound 1 express 17 fantasize 1 fault 1 favor 15 fear 36 feel 35 felt 8 feud 1 figure 32 file 6 find 93 flay 1 flirt 1 focus 1 forbid 4 force 7 forecast 16 foresee 6 forget 6 fret 2 frighten 1 fume 1 gauge 1 get 2 give 7 gloat 1 go 2 grant 1 grip 1 grouse 1 growl 1 grumble 1 guarantee 7 guess 5 gush 1 hail 5 harp 1 have 45 hear 3 herald 1 highlight 2 hint 11 hold 11 hope 136 identify 5 ignore 5 illustrate 7 imagine 3 implore 2 imply 9 impose 4 impress 1 incline 1 include 5 incorporate 1 indicate 76 indict 4 induce 1 influence 1 inform 8 inquire 1 insinuate 1 insist 66 instruct 4 intend 51 interest 4 interject 1 interpret 6 introduce 2 investigate 1 invite 7 involve 2 irk 1 joke 3 jump 1 justify 3 know 128 label 2 lambast 1 lament 3 laud 3 laugh 2 lay 1 learn 13 lecture 1 license 1 like 22 limit 1 link 1 list 6 lobby 1 look 17 love 1 mail 1 maintain 1 make 8 mandate 1 marvel 1 mean 9 measure 1 mention 10 mind 1 misstate 1 moan 1 mount 1 muse 1 name 6 need 2 negotiate 2 nickname 2 nod 1 nominate 1 note 241 notice 4 notify 3 object 2 obligate 1 observe 19 offer 40 oppose 24 opt 1 order 32 outline 5 paint 2 pass 1 pay 1 peg 1 perceive 2 permit 5 persuade 15 pinpoint 2 place 1 plan 175 plead 17 please 1 pledge 8 point 8 poise 1 ponder 3 portray 8 post 76 praise 9 preach 1 predict 101 prefer 11 preoccupy 1 prepare 3 prescribe 2 present 4 press 6 pressure 1 presume 1 pretend 1 prevent 4 prim 1 proclaim 3 produce 1 profess 3 proffer 1 prohibit 13 project 18 promise 34 promote 3 prompt 1 pronounce 2 propose 57 protest 1 prove 3 provide 12 publish 2 purport 1 purr 2 push 5 put 15 quash 1 question 23 quip 6 quote 9 raise 2 rat 2 rave 1 re-emphasize 1 read 8 reaffirm 4 realize 18 reason 6 reassert 1 reassure 2 rebuff 2 recall 44 reckon 6 recognize 14 recommend 30 record 2 recount 4 refer 9 reflect 1 refuse 31 regard 8 regret 2 reignite 1 reiterate 11 reject 13 relate 1 release 2 relieve 1 rely 3 remark 7 remember 4 remind 7 rename 1 renew 1 renounce 1 repeat 3 reply 15 report 313 repute 1 request 14 require 54 resent 2 resist 1 resolve 1 respect 1 respond 16 restate 2 restrain 1 rethink 2 reveal 12 review 1 romance 1 rule 33 ruminate 1 rumor 7 salute 1 satisfy 3 saw 11 say 9017 schedule 1 score 1 scream 1 see 95 seek 32 seem 10 send 1 sense 2 sentence 3 set 4 share 1 shout 3 show 79 shrug 1 sigh 1 sign 1 signal 13 sing 4 snap 2 sniff 4 snort 2 solicit 3 speak 4 specify 12 speculate 24 spell 1 sponsor 1 stagewhisper 1 stand 2 state 19 stipulate 3 stress 20 strive 1 study 4 sue 4 suggest 91 support 14 suppose 14 surprise 2 survey 2 suspect 19 swear 1 take 1 talk 15 teach 1 tear 1 team 1 tell 191 term 4 terrify 1 testify 7 theorize 2 think 333 threaten 14 tout 3 trouble 1 trumpet 1 try 7 turn 1 underestimate 1 underscore 2 understand 19 understate 2 unleash 1 unveil 14 uphold 2 urge 45 value 16 verify 1 view 30

voice 4 volunteer 2 vote 2 vow 18 wad 1 wait 4 want 253 warn 56 welcome 2 whisper 1 wish 9 witness
1 wonder 20 worry 37 wrestle 1 write 38 yell 1

References

- Pareti, S. (2012). A database of attribution relations. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC-2012)*, pages 3213–3217.
- Pareti, S. (2015a). Annotation schema and guidelines for the annotation of PARC 3.0. Technical report, School of Informatics, University of Edinburgh.
- Pareti, S. (2015b). *Attribution: A Computational Approach*. PhD thesis, University of Edinburgh.
- Prasad, R., Dinesh, N., Lee, A., Joshi, A., and Webber, B. (2006). Annotating attribution in the Penn Discourse TreeBank. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, pages 31–38. Association for Computational Linguistics.