

Annotation Guidelines for Experimental Procedures

Developed By
Mohammed Alliheedi
Robert Mercer

Version 1
April 14th, 2018

1- Introduction and background information

What is rhetorical move?

A rhetorical move can be defined as a text fragment that conveys a distinct communicative goal, in other words, a sentence that implies an author's specific purpose to readers.

What are the types of rhetorical moves?

There are several types of rhetorical moves. However, we are interested in 4 rhetorical moves that are common in the method section of a scientific article that follows the Introduction Methods Results and Discussion (IMRaD) structure.

- 1- **Description of a method:** It is concerned with a sentence(s) that describes experimental events (e.g., "Beads with bound proteins were washed six times (for 10 min under rotation at 4°C) with pulldown buffer and proteins harvested in SDS-sample buffer, separated by SDS-PAGE, and analyzed by autoradiography." (Ester & Uetz, 2008)).
- 2- **Appeal to authority:** It is concerned with a sentence(s) that discusses the use of standard methods, protocols, and procedures. There are two types of this move:
 - A reference to a well-established "standard" method (e.g., the use of a method like "PCR" or "electrophoresis").
 - A reference to a method that was previously described in the literature (e.g., "Protein was determined using fluorescamine assay [41]." (Larsen, Frandesn and Treiman, 2001)).
- 3- **Source of materials:** It is concerned with a sentence(s) that lists the source of biological materials that are used in the experiment (e.g., "All microalgal strains used in this study are available at the Elizabeth Aidar Microalgae Culture Collection, Department of Marine Biology, Federal Fluminense University, Brazil." (Larsen, Frandesn and Treiman, 2001)).
- 4- **Background information:** It is concerned with a sentence(s) that deals with method justifications, comments, or observations (e.g., "Unfortunately, our attempts to detect activation of S1P4 expressed in these cells ... were unsuccessful. Therefore, we used CHOK1cells as an alternative host in these studies..." (Holdsworth et al. 2004).

What is a semantic role?

A semantic role is "the underlying relationship that a participant has with the main verb in a clause"¹. For example,

¹ Semantic Role. (2015, December 3). Retrieved August 17, 2017, from <http://www.glossary.sil.org/term/semantic-role>

“An apple was eaten by John”

The sentence describes a frame “eating an apple”, so “John” is the experiencer or agent who eats the apple, and the object “apple” is the patient which is being eaten.

What are the kinds of semantic roles?

There are various semantic roles which already were developed in the literature (e.g., Verb Net).

- 1- **Predicate**: The verb that initiates the frame. It could be a verb or a nominalized verb. Basically, nominalization is “to convert (another part of speech) into a noun, as in changing the adjective *low* into *the lowly* or the verb *legalize* into *legalization*”².
- 2- **Agent**: Initiator of action, capable of volition and most of the times the agent come in phrase like “we” or “the authors” → Proteins were washed three times by **the authors**.
- 3- **Patient**: Affected by action, undergoes change of state → **Haplotypes of the individuals** were reconstructed using BEAGLE (version 4.0) (Browning and Browning 2007).
- 4- **Theme**: Not changed by an action, or being “located” → **Other computing works of this report** were conducted in R (version 2.14.2) (R Core Team 2015), a free software environment for statistical computing and graphics.
- 5- **Instrument**: “used for objects (or forces) that come in contact with an object and cause some change in them. Generally introduced by ‘with’ prepositional phrase”³. Most of the time appears as a Prepositional Phrase (PP).

We have created sub-categories under the semantic role “Instrument” to include:

Note: bold-faced words or phrases are the ones that are referred to in each example.

- (Change) a thing or protocol that can change another thing(s).
 - Note that if the sentence describes selective media which allow only the selected cells to survive while others not. In this case, we should label it as instrument of change.

Example: “Beads with bound proteins were washed six times (for 10 min under rotation at 4°C) **with pulldown buffer** and proteins harvested **in SDS-sample buffer**, separated **by SDS-PAGE**, and analyzed by autoradiography.” (Ester & Uetz, 2008).

² <http://www.dictionary.com/browse/nominalize>

³ Martha Palmer | Projects | Verb Net. (n.d.). Retrieved August 21, 2017, from <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>

- (Measure) a thing or protocol that can measure another thing(s).
Example: “Beads with bound proteins were washed six times (for 10 min under rotation at 4°C) with pulldown buffer and proteins harvested in SDS-sample buffer, separated by SDS-PAGE, and analyzed **by autoradiography.**” (Ester & Uetz, 2008).
- (Observe) a thing which can be used to observe another thing(s)
Example: “The mitochondria was observed **by spinning disk confocal microscopy**”.
- (Maintain) a thing or protocol which can be used to maintain the state of another thing(s).
 You should note that:
 - If the sentence contain inhibition process, you should label it as maintain.
 - If the sentence describe a media that used for growth, you should label it as maintain too.*Example:* “Once the samples were in EPR tubes, they were immediately frozen in liquid nitrogen, and stored **in liquid nitrogen** before using.” (Chen & Guidotti, 2001).
- (Catalyst) a thing that can be used as a catalytic “facilitator” (there are two different types of enzymes, one needs a cofactor to be active as a catalyst and the other doesn’t need a cofactor).
Example: “The ca. 900 bp PCR products were digested **with NdeI and HindIII** and ligated into pUC19.” (Carenbauer et al., 2002)
- (Mathematical) a mathematical or computational instrument (e.g., simulation, algorithm, equation and the use of software)
Example: “Simulations of these EPR spectra were accomplished **with the computer program QPOWA** [30,31]).” (Chen & Guidotti, 2001)
- (Reference) a reference to a paper that describes the complete protocol.
Example: “The preparation of authentic vaccinia H5R protein and recombinant B1R protein kinase were **as previously described [11].**” (Brown et al., 2000)

Other types of semantic roles that occur in some frames:

6- **Goal:** We have categorized into two types:

- a- Physical: A thing that already existed and an action is directed toward it or place to which something moves. → “The ca. 900 bp PCR products were digested with NdeI and HindIII and ligated **into pUC19.**”
- b- Purpose: Used to state author’s intention for doing something. → “**To monitor luciferase cycling**, 0.122 Å— 10 6 cells were seeded per 35-mm plate.”

- 7- **Location:** The physical place where the experiments took place. → “The DNA sequences were analyzed by **the Biosynthesis and Sequencing Facility in the Department of Biological Chemistry at Johns Hopkins University.**”
- 8- **Factitive:** it comes into existence as a result of the event. → **Plasmid libraries** were generated through a two-step cloning process (Kwasnieski et al. 2012 , 2014 ; White et al. 2013).
- 9- Protocol detail:
- 1- **Temperature:** usually comes after the instrument as a prepositional phrase that states the process temperature.
Example: “Beads with bound proteins were washed six times (for 10 min under rotation **at 4°C**) with pulldown buffer and proteins harvested in SDS-sample buffer, separated by SDS-PAGE, and analyzed by autoradiography.” (Ester & Uetz, 2008).
 - 2- **Time (duration):** “class-specific role that is used to express time” (Verb Net project). Usually comes after the instrument as a prepositional phrase that states the process time.
Example: “Beads with bound proteins were washed six times (**for 10 min** under rotation at 4°C) with pulldown buffer and proteins harvested in SDS-sample buffer, separated by SDS-PAGE, and analyzed by autoradiography.” (Ester & Uetz, 2008).
 - 3- **Repetition** of a process:
Example: “Beads with bound proteins were washed **six times** (for 10 min under rotation at 4°C) with pulldown buffer and proteins harvested in SDS-sample buffer, separated by SDS-PAGE, and analyzed by autoradiography.” (Ester & Uetz, 2008).
 - 4- **Condition** of a process or the manner in which it was carried out:
Example: “Beads with bound proteins were washed six times (for 10 min **under rotation** at 4°C) with pulldown buffer and proteins harvested in SDS-sample buffer, separated by SDS-PAGE, and analyzed by autoradiography.” (Ester & Uetz, 2008).
 - 5- **Cofactor:** is classified as “inorganic substances that are required for, or increase the rate of, catalysis.”⁴ Please see the list of most common buffers in Appendix A.
Example:” For phosphorylation, three identical reactions contained H5R protein (70 pmol), B1R protein kinase (90 µl), Tris-HCl, pH 7.4 (20 mM), **magnesium chloride (5 mM), ATP (50 µM), [γ-32P] ATP (50 µCi) and dithiothreitol (2 mM)** in a total volume of 500 µl.” (Brown et al., 2000)
 - 6- **Coenzyme:** is defined as “an organic molecule that is required by certain enzymes to carry out catalysis.”⁴ (In this study, we will call both coenzyme and cofactor a “cofactor”. Cofactor is a hypernym). (Needs to be embedded with the definition above.

4 coenzymes and cofactors. (n.d.). Retrieved September 23, 2017, from http://academic.brooklyn.cuny.edu/biology/bio4fv/page/coenzy_.htm

- 7- **Buffer:** is defined as “a solution containing either a weak acid and a conjugate base or a weak base and a conjugate acid, used to stabilize the pH of a liquid upon dilution.”⁵ Please see the list of most common buffers in Appendix B.

Example:” For phosphorylation, three identical reactions contained H5R protein (70 pmol), B1R protein kinase (90 µl), **Tris-HCl, pH 7.4 (20 mM)**, magnesium chloride (5 mM), ATP (50 µM), [γ -³²P] ATP (50 µCi) and dithiothreitol (2 mM) in a total volume of 500 µl.” (Brown et al., 2000)

- **Examples of annotating semantic roles from our dataset**

- a- “Beads with bound proteins were washed six times (for 10 min under rotation at 4°C) with pulldown buffer and proteins harvested in SDS-sample buffer, separated by SDS-PAGE, and analyzed by autoradiography.” (Ester & Uetz, 2008).

Event 1	Event 2	Event 3	Event 4
Patient: Beads with bound proteins Predicate: were washed Instrument (change): Pulldown buffer Protocol detail: <ul style="list-style-type: none"> - Repetition: six times - Time: 10 min - Condition: under rotation - Temp: 4 C 	Patient: Proteins Predicate: harvested in Instrument (change): SDS-sample buffer	Patient: Proteins Predicate: separated by Instrument (change): SDS-PAGE	Patient: Proteins Predicate: analyzed by Instrument (measure): Autoradiography

- b- “The ca. 900 bp PCR products were digested with NdeI and HindIII and ligated into pUC19.” (Carenbauer et al., 2002)

5 Buffer - Biology-Online Dictionary. (n.d.). Retrieved September 23, 2017, from <http://www.biology-online.org/dictionary/Buffer>

Event 1	Event 2
Patient: The ca. 900 bp PCR products Predicate: were digested Instrument (catalyst): with NdeI and HindIII	Patient: The ca. 900 bp PCR products Predicate: ligated Goal: into pUC19

- c- “The preparation of authentic vaccinia H5R protein and recombinant B1R protein kinase were as previously described [11].” (Brown et al., 2000)

Event 1
Patient: The preparation of authentic vaccinia H5R protein and recombinant B1R protein kinase Instrument (reference): [11] Predicate: described

- d- “A large peak of radioactivity (unreacted ATP) eluted with the water and was discarded, and a smaller broad peak of radioactivity that eluted with 50% acetonitrile was retained and concentrated to 200 µl by rotary evaporation.” (Brown et al., 2000)

Event 1
Patient: a large peak of radioactivity (unreacted ATP) which was already eluted with the water Predicate: discarded
Event 2
Patient: a smaller broad peak of radioactivity which was already eluted with 50% acetonitrile Predicate: retained
Event 3
Patient: a smaller broad peak of radioactivity which was already eluted with 50% acetonitrile Predicate: concentrated Instrument (change): rotary evaporator

- e- “Peptides were sequenced on an Applied Biosystems 476A protein sequencer and phosphorylation sites were analyzed using solid phase Edmann sequencing [20].” (Brown et al., 2000)

Event 1

Patient: peptides

Predicate: sequenced

Instrument (measure): an Applied Biosystems 476A protein sequencer

Event 2

Patient: Phosphorylation sites

Predicate: analyzed

Instrument (measure): using solid phase Edmann sequencing [20]

- f- “Steady-state kinetics constants, K_m and k_{cat} , were determined by fitting initial velocity versus substrate concentration data directly to the Michaelis equation using CurveFit [36].” (Carenbauer et al., 2002)

Event 1

Patient: Steady-state kinetics constants, K_m and k_{cat}

Predicate: determined

Instrument (mathematical): by fitting initial velocity versus substrate concentration data directly to the Michaelis equation using CurveFit [36]

2- Annotation guidelines

These guidelines describe a classification scheme for the Method section in biochemistry articles which are concerned with the rhetorical moves and semantic roles.

1- Before the annotation

The annotator should read the entire article. This is very important as we are only looking for the annotation of one section in biochemistry articles (i.e., the Method section). Thus, the interpretation of some sentences in the Method section can become clear once the entire article has been read. Please note that you don't need to understand the article in detail, you can go back and forth between sections in the article. Please also try to focus on the main four rhetorical moves and ensure that the sentence is concerned with one of these moves.

2- During annotation

Annotation should be proceed by only annotating one sentence at a time and assigning this sentence to one of the moves. Usually consecutive sentences are marked with the same move type. You can label consecutive sentences with the same move if these sentences share the same move. First you should find the verbs in every sentence, and then the annotator should be able to answer the following questions:

Q1. Can you identify the predicate (e.g., verb, nominalized verb, adjective) in the sentence?

If yes, you should label it as "predicate". Then go to Q2

Q2. Can you identify the patient, theme, or factitive from this sentence or phrase? If yes, you should label it as either patient, theme or factitive depends on their definition giving in (section 1), and then proceed to Q3.

Q3. Can you identify the instrument in this sentence or phrase?

If yes, you should label it as "instrument" and select one of the instrument types (e.g., Change, Measure, Observe, Maintain, Catalyst, Mathematical and Reference) then proceed to Q4. If not, you can proceed to Q4

Q4. Is there additional information in the sentence such as process temperature, time, or buffer used in the experiment?

If yes, you should list this information under protocol detail and proceed to Q5. If no, you should proceed to Q5.

Q5. Can you identify a goal either (**physical**: where the theme, patient, or factitive in this sentence is directed to OR **purpose**: which indicates the author intention in the sentence)?

If yes, you should label it as "Goal: physical" for the first type or as "Goal: purpose" for the second one, and proceed to Q6. If no, you should proceed to Q6.

Q6. Can you identify the location where the experimental process took place?

If yes, you should label it as “location” and then proceed to Q7. If no, you should proceed to Q7.

Q7. Does this sentence describe an experimental procedure?

If yes, you should label it as “**description of the method**” and proceed to next sentence in the paragraph. If no, you should proceed to Q8.

Q8. Does this sentence use a technique, protocol or method that was previously introduced in the scientific field?

If yes, you should label it as “**appeal to authority**” and proceed to next sentence in the paragraph. If no, you should proceed to Q9.

Q9. Does this sentence talk about method justifications, comments, or observations?

If yes, you should label it with “**background information**” and proceed to next sentence in the paragraph. If no, you should proceed to Q10.

Q10. Does this sentence list or describe experimental materials?

If yes, you should label it as “**source of the materials**” and proceed to examine the next sentence. If no, you should proceed to next sentence in the paragraph.

Appendix A:

A List of most common buffers⁶:

Buffer

MES

Bis-Tris

ADA

ACES

PIPES

MOPSO

Bis-Tris Propane

BES

MOPS

TES

HEPES

DIPSO

MOBS

TAPSO

Trizma

HEPPSO

POPSO

TEA

EPPS

Tricine

Gly-Gly

Bicine

HEPBS

TAPS

AMPD

TABS

AMPSO

CHES

CAPSO

AMP

CAPS

CABS

⁶ <http://www.sigmaaldrich.com/life-science/core-bioreagents/biological-buffers/learning-center/buffer-reference-center.html>

Appendix B:

A List of most common cofactors⁷:

Cofactor
Thiamine pyrophosphate ^[29]
NAD⁺ and NADP⁺ ^[30]
Pyridoxal phosphate ^[31]
Methylcobalamin ^[32]
Cobalamine ^[5]
Biotin ^[33]
Coenzyme A ^[34]
Tetrahydrofolic acid ^[35]
Menaquinone ^[36]
Ascorbic acid ^[37]
Flavin mononucleotide ^[38]
Flavin adenine dinucleotide ^[38]
Coenzyme F420 ^[39]
Adenosine triphosphate ^[40]
S-Adenosyl methionine ^[41]
Coenzyme B ^[42]
Coenzyme M ^{[43][44]}
Coenzyme Q ^[45]
Cytidine triphosphate ^[46]
Glutathione ^{[47][48]}
Heme ^[49]
Lipoamide ^[5]
Methanofuran ^[50]
Molybdopterin ^{[51][52]}
Nucleotide sugars ^[53]
3'-Phosphoadenosine-5'-phosphosulfate ^[54]
Pyrroloquinoline quinone ^[55]
Tetrahydrobiopterin ^[56]
Tetrahydromethanopterin ^[57]