

Annotation Guidelines

<Revised July 20, 2005>

HAGIT SHATKAY¹, W. JOHN WILBUR², AND ANDREY RZHETSKY³

1. School of Computing, Queen's University, Kingston, ON, Canada (shatkay@cs.queensu.ca)
2. National Center for Biotechnology Information NLM, NIH, Bethesda, MD, USA (wilbur@ncbi.nlm.nih.gov)
3. Department of Biomedical Informatics, Columbia University, New York, NY, USA (ar345@columbia.edu)

Overview

Our general aim is to identify information-bearing fragments (*high-utility fragments*) within scientific text, in order to substantiate our knowledge about important biomedical entities and processes. Furthermore, we would like to differentiate these high-utility fragments from non-informative ones automatically, and to distinguish among several types of high-utility fragments.

We are pursuing two related sub-goals: 1) to manually annotate a sizable corpus, and 2) to use this corpus to build and train text-classifiers.

To approach the first subtask, annotation of a biomedical corpus, we characterize text fragments along the following dimensions:

- **Focus.** We distinguish several types of contents that scientific assertions convey: *scientific* findings, the experimental *methodology* used, and *generic* issues, such as the organization of the paper or the state of science.

We acknowledge that the focus of an assertion is context-dependent. That is, the focus of the same statement may be viewed differently, depending on the specific paper, and on the section in which it appears. Typically, however, the paragraph, and even the sentence, within which the assertion occurs are sufficient to provide context for making an informed decision about its focus. This point is further discussed in the detailed guidelines below.

- **Polarity:** An assertion can be either *positive* (*P*) or *negative* (*N*).
- **Certainty:** Indicates the strength of the author's belief regarding the validity of the assertion.
- **Evidence:** Indicates whether the assertion is supported by evidence, and distinguishes between several types of supporting evidence.

As was the case for Focus, the perceived level of evidence can be context-dependent. Some statements may be viewed differently as containing or lacking evidence – depending on the scope of the text in which they occur. For instance, within a sentence discussing experimental results that were conducted and observed by the authors, statements of experimental findings may implicitly reference the experiment or the results using pronouns ("*it* shows...", "*these* results suggest..."), and as such are viewed as *supported by evidence*. However, similar statements may be made within sentences in review papers, where the authors' own experiments and results are not presented. In such cases, explicit citations or references to other experimental papers are needed, and their occurrence *is considered as evidence* to support the claims. (See examples in *Appendix E*).

This point is further discussed in detail in the guidelines below.

- **Direction/Trend:** Indicates whether the assertion reports an *increase* or a *decrease* of a quantity/quality, or provides a quantitative statement of *high/low levels* of a quantity/quality with respect to a specific phenomenon, finding or activity.

Unlike Focus and Evidence, which are context-dependent, the three other dimensions: Polarity, Certainty, and Direction/Trend, denote *local properties*. That is, every individual statement can be independently evaluated for Polarity, Certainty and Direction/Trend, regardless of the broader context in which it occurs.

We next define the annotation guidelines and discuss the above dimensions in more detail.

Examples demonstrating the use of these guidelines are provided in the appendices to this paper. These examples are an essential part of these guidelines, and should be carefully read and understood prior to the beginning of the annotation. They should also be consulted when there is doubt about borderline cases.

Annotation

Tagging Granularity (paragraphs, sentences, or sentence-fragments)

Before providing the annotation scheme itself, we define the annotation unit, i.e., the scope of a text-unit that is being assigned a tag. Paragraphs (and, of course, coarser constituents – such as sections) are typically too heterogeneous in their scientific contents, focus, evidence, and polarity to enable a single tag assignment. An exception is within the “methods” section – where most paragraphs make a positive statement describing a single action or action-sequence, or a single used method. In most other sections, sentences vary greatly in scientific content and polarity. Even within sentences, there is often variability in content, polarity, and the level of evidence.

Therefore we require a separate tag (actually – it is a “tag-sequence”) for each *fragment within a sentence* – for which one can assign categories of interest (as described below). Under this scheme, each annotator decides independently the boundaries of the annotated fragments.

Fragmentation should occur at each point in the sentence where the annotator perceives a change in focus, polarity, certainty or any of the other attributes that we define here. *Fragmentation should be limited to these cases, and should not be performed otherwise.* Examples for fragmentation of sentences are provided in *Appendix A*.

Tagging the fragments:

Each fragment will be tagged with a sequence (an ordered list) of letters and numbers, denoting several types of information, as follows:

Fragment Number: The ordinal number of the fragment within the sentence.

The tag of the first fragment within a sentence starts with the number 1, the second with 2 etc. The tag of an un-fragmented sentence always starts with 1. (*See examples in Appendix A*).

Focus: Each text fragment may convey one (and sometimes more) of:

- **Scientific** content, findings and discovery; we refer to this type of information as *Science*, and indicate it by the tag *S*.
- **Generic-level** information: General state of knowledge and science outside the scope of the paper, the structure of the paper itself or the state of the world. Such statements are not usually based on scientific experiment, and may reflect an opinion or an observation that would have been as truthful, and probably as valid, if made by a layperson. We refer to it as *Generic*, and denote it with the letter *G*.
Examples of Generic fragments are provided in Appendix C.
- **Methodology** that was used to execute an experiment or a study. We refer to it as *Methodology*, and denote it with the letter *M*. *Examples of Methodology fragments are provided in Appendix D.*

In some cases a fragment discussing Methodology may also discuss Science. In such cases the leftmost tag-letter will be *M* followed by *S*. Analogously, we also allow for other combinations of tags: *GM*, *GS*, *GMS*.

We advise though that the compound tags should be used sparingly – only when the annotator is unable to assign a narrower category to the fragment.

The focus of an assertion is context-dependent. As stated earlier, the focus of a statement may be viewed differently depending on the context (paper, section, paragraph and sentence) in which it appears. What may be regarded as a *scientific finding* in one context is a *methodology* in another. In fact, most scientific methods are based on what were at one time reported scientific findings. They began as a sequence of experimental preparations followed by observations. Later they became named procedures, where the former “observations” have come to be regarded as consequences and are taken for granted. For instance, the ability to use bacterial artificial chromosomes (BAC) for DNA replication was a finding when first introduced in papers discussing cloning as a new technology; however, in most current papers discussing genomics, BACs are mentioned as standard methodology.

Thus, one may encounter accounts describing sequences of preparations and observations that can be challenging to separate into methodology vs. scientific findings. We therefore expect that annotators will take into account the specific context in which a statement is being made, to decide if a statement indicates a methodology, a scientific finding, or both. *See Appendices B-F for annotated examples.*

Focus summary: *The leftmost letter, which follows the fragment number in the tag-sequence, is either S (Science), G (Generic), or M (Methodology). Rarely may it be MS, GM, GS, or GMS.*

Polarity: A fragment with any focus can be stated either positively or negatively. The next field in the tag-sequence is thus either a ***P (positive)*** or an ***N (negative)***.

For statements that convey lack-of-knowledge, (e.g. “*It is still unknown whether...*”), the default assignment is ***P***. The lack of knowledge in this case will be reflected with a certainty degree of ***0***, as explained in the next item.

Every fragment should be annotated by its polarity, regardless of its focus.

Certainty: Each fragment conveys a degree of certainty about the validity of the assertion it makes. The annotation will use ***a scale of 0-3 as a measure of certainty***, for both positive and negative statements.

The lowest degree (*0*) represents *complete uncertainty*, that is, the fragment explicitly states that there is an uncertainty or lack of knowledge about a particular phenomenon (“*it is unknown if...*” or “*it is unclear whether...*” etc.).

The highest degree, (*3*), represents complete certainty (an accepted, known and/or proven fact).

Intermediate *degrees*: (*1*) represents a low certainty, while (*2*) is assigned to *high-likelihood* expressions that are still short of complete certainty.

When assigning tags, the annotator should not downgrade the certainty rating based on his/her own impression of the statement’s inaccuracy. That is, as a rule of thumb, if a statement *does not* explicitly indicate any degree of *uncertainty* – it should be taken to convey complete certainty, and annotated with certainty level 3. Statements should be tagged as uncertain if and only if the text itself contains language that supports this impression.

Our goal is to infer the *certainty intended by the author*. Therefore, the tag should reflect the degree of certainty conveyed by the author’s statement about the scientific finding, rather than the annotator’s belief.

Every fragment should be annotated by a degree of certainty, regardless of its focus.

Evidence: Indicates, for a fragment, regardless of its focus and certainty, if its assertion is supported by evidence. The existence – or the lack of –evidence is denoted in the tag starting with the letter *E*. The letter is followed by one or more digits, in the range [0-3], indicating the type of evidence or its absence thereof:

- *E0*: There is no indication of evidence in the fragment whatsoever, or an explicit statement is made in the text about the *lack of evidence*.
- *E1*: There is a claim of evidence, but no verifying information is explicitly given. Evidence is not shown within the annotated sentence/fragment, and no explicit reference to it is provided. The evidence is merely asserted to exist in some form, possibly in the preceding text, or in prior experiments. The location of the evidence is not explicitly stated but only eluded to. (That is, it is not clear which part of prior text is relevant, and there is no reference to the specific literature, experiments or figures.) Note that in this case the indirect implication of evidence may not be explicit in the fragment, but implied by a use of term referring to a previous fragment. For instance, a sentence may begin with the fragment “*Previous experiments show that...*”, followed by the fragment, “*therefore, it is likely that ...*”. Both fragments are of evidence level 1; the first because it points to experiments without an explicit reference, and the second, because of the “*therefore*” term which uses the previous assertion as an indirect evidence.
- *E2*: Evidence is not given within the sentence/fragment, but *explicit* reference is made to other *papers* (citations) to support the assertion.
- *E3*: Evidence is provided, within the fragment, in one of the following forms:
 - a. A reference to experiments previously reported within the body of the paper by a direct description of the finding as an experimental result (e.g. “Our data indicates...”, “...our results show”...)
 - b. A verb (typically in the past-tense) within the statement indicates an observation or an experimental finding which is described within the paper, (e.g. “We found that...”, “We see that...”, “The level of ... increased over time...”).
 - c. A reference to an experimental figure or a table of data given within the paper.

We note that determining the level of evidence may depend on the way in which past experience or other parts of the paper are referenced within the annotated fragment:

First, not every reference within a fragment is “evidence”. Some references are to figures that are merely illustrations or clarifications rather than experimental results. It is therefore important not to mechanically annotate each reference in a fragment as “evidence” of type *E3*. A careful reading of the sentence/fragment can help to distinguish a reference to evidence from a reference to an illustration.

As a rule of thumb: *When it isn't clear that a reference indeed points to evidence, don't use the E3 code.*

Second, a statement about a certain finding may be assigned different levels of evidence depending on the wording used in the statement. For instance, something reported as a finding by the authors would be annotated as *E3*. (e.g., “*Our data demonstrate that ICG-001 has no effect on AP1 and CRE reporter constructs*”). In this case the words “*Our data demonstrate*” is the evidence. However, a similar statement may occur without an indication that an experiment was done. (e.g., “*ICG-001 has no effect on AP1 and CRE reporter constructs*”). In such a case it would be annotated as *E0* – if it is stated without any support. This same statement would be annotated as *E1* if accompanied by a non-explicit reference (e.g. “*Previous studies suggest that ICG-001 has no effect on AP1 and CRE reporter constructs*”). Finally, if explicit reference to the original work is given: “*Previous studies suggest that ICG-001 has no effect on AP1 and CRE reporter construct (25)*”, the tag would be *E2*. The important point here is that it is not the statement of scientific details themselves, be they ever so intricate, that constitutes the evidence. Rather it is the indication in the wording that gives some definite direction

about where the proof of the statement could be found that constitutes the evidence that is to be coded in the annotation.

In all these cases, it is imperative that the annotator *carefully consider the context and the subtle use of words* to distinguish between a report by the authors of their own observation, (which is essentially a scientific evidence), and any other positive statement describing a scientific fact which – while it may be true – is unsubstantiated by evidence.

Please refer to the attached appendices for numerous examples demonstrating these distinctions.

At times a fragment may contain more than one type of evidence. For example, the fragment: “...the overexpression of phospho-H2Av did not induce G2/M arrest or affect DSB-dependent G2/M arrest (fig. S10) (14,21).” contains indication for evidence both in a figure (fig. S10) within the paper (**E3**) and in other cited sources (14,21) outside the paper (**E2**). In such a case, all applicable evidence codes should be indicated as tags, following one another. “**E23**”.

On the other hand, when breaking a sentence into fragments, the evidence may be associated with only one of the fragments, leaving another fragment without any indication of evidence. In this case – the evidence tag **E0** should be assigned to the fragment lacking any indication of evidence.

(For example: “Our experiments demonstrate that ABC is highly expressed in ..., but not in...”

The first fragment: “Our experiments demonstrate that ABC is highly expressed in...” is tagged as **SP3E3+**, while the second fragment “but not in...” is tagged as **SN3E0** (No evidence. There is also no direction/trend in this case). The reason for the fragmentation is the change in polarity from positive to negative.)

Direction/Trend: The last (rightmost) part of the tag-vector for a fragment indicates whether the assertion reports a qualitatively *high/low* level or an *increase/decrease* in a specific phenomenon, finding or activity. An increase or high-level is denoted by + while a decrease or low is denoted by -.

This tag is introduced to separate the notion of positive/negative results and assertions (as captured by *Polarity*) from the level of the observed phenomenon itself.

For instance, the sentence: “In fact, as demonstrated using several SOD assays including pulse radiolysis, 2-ME does not inhibit SOD” indicates a negative experimental finding (“did not...” – negative polarity), about a negative trend (“inhibit”). This is a case known as *double-negation*, and is typically hard to annotate, as it is not clear whether the phenomenon is actually present or not.

Separating *Direction* from *Polarity* provides a mechanical way to annotate and interpret the statement. Moreover, this separation also provides means to indicate presence/absence of experimental findings (tagged as *Polarity*), regardless of whether these findings demonstrate the presence or the absence of the monitored phenomenon (the latter is captured by the *Trend*).

Every indication of trend occurring in the text should be annotated.

Annotation summary:

The typical fragment annotation thus consists of a tag of the form:

[<Integer>][S|G|M]⁺[P|N][0-3](E[0-3]⁺)[-/+][?]

where <Integer> is the ordinal number of the fragment within its sentence, starting at 1.

That is: Fragment number, followed by *Focus* indicator(s), then by *Polarity*, then by degree of *Certainty*, then by an *Evidence* code(s), and *optionally* ended with a *Direction* indicator.

Appendix A

Fragmentation examples. Where to break sentences. Sentences that remain un-fragmented.

Appendix B

Examples of many well-behaved annotations – mostly science focus, some others

Appendix C

Examples of Generic statements

Appendix D

Examples of Methodology statements

Appendix E

Examples of the various types of Evidence in statements

Appendix F

Examples of interesting border-line cases in polarity, direction, methodology - and other basket cases ☺