

THYME Annotation Guidelines

Developed by Will Styler, Guergana Savova, Martha Palmer, James Pustejovsky, Tim O’Gorman, and Piet C. de Groen, based on the ISO TimeML temporal relations specification

Last updated: February 9, 2014

Contents

1	Introduction and Philosophical Notes	4
1.1	Introduction	4
1.2	Acknowledgements	4
1.3	About the examples in this guide	5
2	Overview of the annotation process	5
2.1	Event Annotation – Finding and annotating timeline-relevant events and states	5
2.2	Relation Annotation – Temporally linking events and timepoints together	6
I	Event Annotation Stage	6
3	EVENT Annotation	6
3.1	Overview	6
3.2	Finding Events	7
3.2.1	Deciding what is timeline-relevant	7
3.2.2	Finding Syntactic Heads	8
3.2.3	Some Exceptions and Specific Patterns	11
3.2.4	Specific Prohibitions	14
3.3	Annotating DocTimeRel of EVENTS	15
3.3.1	BEFORE	15
3.3.2	OVERLAP	16
3.3.3	AFTER	16
3.3.4	BEFORE-OVERLAP	16
3.4	Annotating type of EVENTS	17
3.4.1	N/A	17
3.4.2	ASPECTUAL	17
3.4.3	EVIDENTIAL	18
3.5	Annotating polarity of EVENTS	18
3.5.1	POS	18
3.5.2	NEG	19
3.6	Annotating degree of EVENTS	19
3.6.1	N/A, MOST and LITTLE	20
3.7	Annotating contextual modality of EVENTS	20
3.7.1	ACTUAL	20
3.7.2	HEDGED	20
3.7.3	HYPOTHETICAL	21
3.7.4	GENERIC	22
3.8	Annotating contextual aspect of EVENTS	23
3.8.1	N/A	23
3.8.2	NOVEL	23
3.8.3	INTERMITTENT	23
3.9	Annotating permanence of EVENTS	24
3.9.1	FINITE, PERMANENT and UNDETERMINED	24

4	TIMEX3 Annotation	24
4.1	Identifying and Annotating TIMEX3s	24
4.2	Annotating TIMEX3 class	25
4.2.1	DATE	25
4.2.2	TIME	26
4.2.3	DURATION	27
4.2.4	QUANTIFIER	27
4.2.5	PREPOSTEXP	28
4.2.6	SET	28
II	Relations Annotation Stage	28
5	Narrative Containers	29
5.1	Why Narrative Containers?	29
5.2	Narrative Containers from DocTimeRel	30
5.3	Annotating additional narrative containers	31
5.4	EVENTs as Narrative Container Anchors	32
5.5	Single-bounded narrative containers	32
5.6	Nested Narrative Containers	32
5.7	Choosing the Anchors of Narrative Containers	34
5.8	Ordering within narrative containers	35
6	TLINK Annotation	35
6.1	TLINK sub-types	35
6.1.1	BEFORE	35
6.1.2	CONTAINS	36
6.1.3	OVERLAP	37
6.1.4	BEGINS-ON	38
6.1.5	ENDS-ON	38
6.2	When to TLINK	38
6.2.1	TLINK only when it captures more information than just marking DocTimeRel	39
6.2.2	TLINK all EVENTs to their narrative container, if possible	39
6.2.3	TLINK all explicitly stated temporal relations	39
6.2.4	Try to only link EVENTs and TIMEX3s within the same sentence	40
6.2.5	ACTUAL or HEDGED EVENTs should never be linked to HYPOTHETICAL or GENERIC EVENTs, and vice versa.	40
6.2.6	You do not need to TLINK TIMEX3s to one another	40
6.3	Expressing TLINK Types in Point Algebra	40
6.4	Tricky TLINK examples	41
7	ALINK Annotation	41
7.1	ALINK sub-types	42
7.1.1	CONTINUES	42
7.1.2	INITIATES	42
7.1.3	REINITIATES	42
7.1.4	TERMINATES	43
8	DOCTIME and SECTIONTIME Annotation	43
9	Annotating Pathology Notes	44
9.1	AJCC Staging Codes	45
9.2	Guidelines for Annotating Pathology Notes	45
10	Schema Description	47

10.1 Default Values	48
11 Special Cases and Constructions	48
11.1 Avoiding Inference	48
11.2 CONTAINS and NEGATIVE Polarity EVENTS	49
11.3 Discussions, Generalizations, and other EVENTS which aren't on the timeline	49
11.3.1 Ground Rules for annotating verbs of discussion	50
11.3.2 Examples of discussion	50
11.4 Family History/History of...	51
11.4.1 Medical History sections	52
11.5 Pre- and Post- expressions (preoperative, post-treatment, etc)	52
11.6 Annotating the word "plan" or "plans"	53
11.7 Annotating "prior" and "at the same time"	54
11.7.1 Annotating "Prior"	54
11.7.2 Annotating "at the same time"	55
12 Non-annotated sections of clinical notes	55
12.1 Duplicated Text	56
13 Glossary	57
14 References	58
15 Document Revision History (A Brief History of THYME)	59
16 Appendix 1: Clinical Note Section Labels	62

NOT FOR REDISTRIBUTION

1 Introduction and Philosophical Notes

"Everyone's life consists of dates, giving life a rhythm and sense of gradation, as if from the eminence of a date one could look back and down, and see the past itself. A clear, comprehensible past, divided up into squares of events, lines of paths taken." - Andrey Kurkov ¹

1.1 Introduction

Thank you for becoming an annotator! The THYME project (short for "Temporal Histories of Your Medical Events") aims to develop a human-annotated corpus of medical records in which the temporal relations (or relations in time) between different events, occurrences, states, dates, and procedures are clearly annotated. This corpus will then be used to aid in machine learning, in hopes that this annotation can then be passed onto a machine such that the more than 40,000,000 records in the Mayo Clinic electronic medical records (EMRs) (and those of other institutions) can be searched both by doctors and by researchers hoping to find long-span correlations and tracking patient outcomes.

Please note that we will only be working with medical records that have been de-identified. This means that no references to the patient's name or birthdate will be included, the names of different doctors will be obscured, and that all dates have been modified from their original form. However, because we're examining temporal relations, all relations between the dates have been preserved. So, to give an example, every date for a given patient may have had 16 days added to it. The number added (or sometimes subtracted) is a random number which is kept consistent across each individual patient and record, but is not revealed to anybody but the clinic personnel preparing the documents, and changed between patients. Even with this de-identification done, remember, this is still actual patient data, and your discretion and caution with the data is necessary.

Your annotation will consist of two main tasks. First, you will search the document and annotate events relevant to the patient's treatment and life, and provide additional temporal and modal information about these events. Then, you'll find discrete references to times and dates which we can use as pins on the grand timeline of the patient's history. Then, finally, you'll go through and mark the temporal relations of these events to one another. By doing this, you will be generating a useful timeline of the patient's medical care which is temporally rich enough for any number of uses.

The final goal is to produce annotations rich enough that a computer, using complex inferencing, co-reference, and domain-specific algorithms, would be able to construct an accurate timeline of when the events in a given medical record occur relative to any fixed dates present and relative to one another. This timeline could then be viewed directly by physicians reviewing a patient's case history, or queried by means of natural language questions ("Did the patient report any bleeding before her diagnosis of Colon Cancer?"). Such a tool could be very useful for increasing both the speed with which a doctor can "get current" with a patient's status, and potentially, increase the power of statistical queries for running analyses of large banks of medical data to find patterns too temporally distributed for humans to notice.

1.2 Acknowledgements

The THYME annotation schema you'll be using was developed in a series of steps by a joint task force of people from the University of Colorado, the Mayo Clinic, and the Harvard Medical School/Boston Children's Hospital. It was originally based on the TimeML schema annotation guidelines created by Roser Sauri, Jessica Littman, Bob Knippen, Robert Gaizauskas, Andrea Setzer, and James Pustejovsky (<http://timeml.org>) and the ensuing ISO Standard. The first iteration of the schema was developed by Steven Bethard, Will Styler, Donna Ihrke, Martha Palmer, and Guergana Savova and described in Savova et al. (2009) *Towards temporal relation discovery from the clinical narrative* [3]. The subsequent major revision of the schema (bringing it to the state described in this document) was done by Will Styler, Martha Palmer, and Guergana Savova with heavy input from James

¹Kurkov, Andrey. *Death and the Penguin (Melville International Crime)* (p. 141). Random House Inc Clients.

Pustejovsky, Steve Bethard, Piet de Groen, Dr. William F. Styler III, David Harris, Glenn Zaramba, Jiaping Zheng, Margarita Sordo, and Jim Martin.

The project described was supported by Award Number R01LM010090 from the National Library Of Medicine. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Library Of Medicine or the National Institutes of Health.

1.3 About the examples in this guide

In our examples, words in [brackets] are to be annotated with the type of annotation or attribute under discussion. The brackets represent the proper span. Spans of text in {Curly Brackets} represent TIMEX3 annotations which are marked, even if not being discussed explicitly in the example.

In these guidelines, TLINKs will be represented with the below format:

- (1) Item B comes before Item A
 - a. (Item A) BEFORE (Item B)

Where the first line is the actual sentence, Item A is the first EVENT or TIMEX3, Item B is the second EVENT or TIMEX3, and BEFORE (or CONTAINS, OVERLAP, etc) is the TYPE of the TLINK, described in Section 6.1.

2 Overview of the annotation process

2.1 Event Annotation – Finding and annotating timeline-relevant events and states

First, you'll need to go through the document and find any events or states relevant to the patient's clinical timeline. This could range from procedures to diseases to diagnoses to patient complaints and states, as well as nearly anything else relevant to the patient's case. Each of these clinically relevant events will be given an *EVENT* annotation.

Then, for each *EVENT* annotation, you'll specify further attributes. These attributes specify the temporal relation of the event to the time of service (called "*DocTimeRel*", which specifies whether an *EVENT* happened before, after, during, or before and during the patient's visit), whether the event really happened or whether it was hypothetical, negation, its aspectual type, and more. This will be done for each event at the same time that you create the annotation.

In some datasets, you may have *EVENTs* pre-marked for you (based on data from other layers of annotation). For those pre-marked *EVENTs*, your first step will be to go through and update the given *EVENTs* with the proper attributes, as described above, and to remove any annotations that are given to you, but aren't actually *EVENTs*.

Once you've found all the events in the note, you'll go back through and look for discrete references to time (called "*TIMEX3s*" in this schema). These can be dates (like "June 9th" or "May 2006"), times ("5:15pm"), relative times ("In six months" or "Next week"), or durations ("For 15 years", "Since last year"). These will be marked as *TIMEX3* annotations, and then marked for their types. In addition, you'll find and mark *DOCTIME* and *SECTIONTIME*, where applicable, and then, you'll send your data off for what we refer to as "entities adjudication". In this process, your annotations will be compared to those of the other annotator working on the same set, and a consensus will be reached, giving us a "gold standard" set of *EVENT*, *TIMEX3*, *DOCTIME* and *SECTIONTIME* annotations.

2.2 Relation Annotation – Temporally linking events and timepoints together

Once the events and timepoints have all been reviewed, vetted, and solidified by an adjudicator, your next step is to mark the relations of these EVENTS to other EVENTS and TIMEX3s by creating a series of *temporal link* ("TLINK" for short) annotations. TLINKs connect two EVENTS, or an EVENT and a TIMEX3 together, specifying the temporal relationship between them (before, overlap, contains, begins-on and ends-on). So, if one event occurred before another, you would link them with a BEFORE TLINK. If one event started on a certain TIMEX3, you would link them with a TLINK of the "BEGINS-ON" type. Armed with these types of TLINK, you'll go through the entire document and mark the temporal relations between EVENTS and TIMEX3s within each section. When doing this, we'll make strong reference to the idea of "narrative containers", the sorts of temporal "buckets" that many EVENTS fall into, whether they be dates, other EVENTS, or even abstract concepts like "family history".

This placing of EVENTS within these narrative containers is a very important step, as every event needs to be anchored to the timeline somehow. Whether this anchoring is as specific as a TLINK to a TIMEX3 or general as the DocTimeRel marking (which temporally links the EVENT to the time of the document's writing), there needs to be something. An EVENT that's just floating with no DocTimeRel or TLINK isn't very useful, and in general, we need to make sure that no event is left stranded and completely unable to attach itself to the timeline. When in doubt, make the annotation, whether it be an EVENT, a TLINK, or an attribute. Redundancy will never hurt.

Once you've finished TLINKing the document, you'll want to make sure that any aspectual EVENTS (like "continues" or "restarted") are linked to the EVENT they modify using *aspectual links* ("ALINKs").

Part I

Event Annotation Stage

The first stage in THYME annotation is to go through the document and mark EVENTS, TIMEX3s which are relevant on the clinical timeline, as well as DOCTIME and SECTIONTIMEs. Each EVENT or TIMEX3 marked will also be annotated for particular characteristics which give us more information and detail about each item. Essentially you are defining the points in our timeline, before we go back and actually mark how they relate to each other temporally.

3 EVENT Annotation

3.1 Overview

As mentioned above, one of the key steps in THYME annotation is finding the events in the document which are relevant to the clinical timeline, and then marking their properties.

This is complicated significantly by the fact that you will only be annotating a single word for each event - that event's SYNTACTIC HEAD. Thus in section 3.2 we will explain why the following sentence has two things we will consider timeline-relevant, "the newest MRI" and "a previously undiscovered mass":

(2) <The newest MRI> revealed <a previously undiscovered mass>.

What you will actually annotate within this sentence, however, are the heads of these two phrases as defined in section 3.2.2. As such, you will actually end up labeling just two words, "MRI" and "mass":

(3) The newest [MRI] revealed a previously undiscovered [mass].

These two words will be annotated as events in the annotation tool, and it is these elements which will be annotated with the semantic characteristics discussed in sections 3.3- 3.9 .

3.2 Finding Events

What is an EVENT? In our schema, an EVENT is anything that's relevant on the clinical timeline. Put differently, anything that would show up on a detailed timeline of the patient's care or life would be considered an EVENT. So, a diagnosis would certainly appear on such a timeline, as would a tumor, illness, or procedure, but temporally span-less entities like people (the patient's mother-in-law or the doctor), organizations (the emergency room), or non-anatomical objects (the patient's car) will never be EVENTS.

Another way to think about this is that if it wouldn't make any sense to relate a given span of text to the document time, chances are, the span doesn't represent an EVENT. So, mention of a patient's arm is not an EVENT, but the same arm swollen (or missing) is most certainly an EVENT, as these states have certainly changed throughout the patient's life and treatment. Similarly, it makes a great deal of sense to relate an MRI, for instance, to the document time ("The MRI occurred AFTER the document was written"). So, in this example, MRI would be an EVENT, but the patient's arm would not.

In this schema², EVENT items don't necessarily have to be actual events in the sense in which the word is conventionally used. "Event" is essentially any structure relevant to the timeline, and therefore states and conditions can be EVENTS just as easily as surgeries can. This also means that EVENTS don't have to be verbs either – adjectives and nouns often be marked as EVENTS as well, such as "the eye is [swollen]"

Note also that although we are putting these on a timeline, you can label many EVENT entries which refer to the same actual "event" on a timeline. You can (and should) label the same heart attack as an EVENT ten different times in the document, so long as it's a separate event in each. We have a secondary "coreference" annotation project which will figure out which events are the same, and therefore we want to capture every EVENT regardless of its previous mentions.

3.2.1 Deciding what is timeline-relevant

This first thing one needs to do is to decide which stretches of text seem to represent things which will go on the timeline. The general approach to that is simple – verbs (even when intimately tied to their arguments) should be thought of using only the verb itself, since our Propbank project has complex models of verbal argument structure anyways. Noun phrases and adjective phrases, in contrast, should be looked at as the full units of meaning. In the end we will only tag a single word for each event, but it's useful to think first about the whole span of that what that even encompasses.

Furthermore, although it might be tempting to label verbs as EVENTS, we actually try to never annotate "semantically light" predications when they function only to comment on or introduce the (usually nominal or adjectival) event that is actually relevant. For example, in sentences like (4) below, the state of being usual is not, in itself, relevant to a clinical timeline, and is therefore not be labeled as an EVENT- we would instead only be looking at "colonic colon cancer" and "metastasize to the distal esophagus" as events for the timeline:

(4) It would be unusual for <colonic colon cancer> to <metastasize to the distal esophagus>

This approach applies to all predications of necessity, possibility, expectation and so forth. Naturally we also ignore "copular" verbs when the more eventive, important part is the noun or adjective they modify. One might therefore note that "is" in the following phrases is not annotated, but rather the entities which follow them, "a dilated esophagus" and "stable":

(5) Given these considerations, this is likely just <a dilated esophagus>.

²Note that in the TimeML ISO standard (ISO 24617-1:2009(E):2009(E)), EVENTS such as those discussed here are more broadly referred to as "eventualities". "event" has a more specific meaning in ISO TimeML.

- (6) The patient is <stable> at {this time}

This idea of "semantically light" also applies to verbs of having and experiencing, so that the following phrases would only look at "nausea", "peripheral edema", and "neck pain", and not the verbs introducing them:

- (7) The patient has <nausea> and <peripheral edema>
 (8) She has been experiencing <neck pain> since {July}

Similarly, verbs such as "perform", "develop" and "undergo" are also ignored, as they similarly do not add a separate timeline point but simply introduce the event or condition of their object:

- (9) We will perform <a hemicolectomy>
 (10) {April 2005} - developed <constipation> for {several days}

A second major criteria for eventiveness is legal relevance. Medical records do not exist solely to capture medical information, but are also legal documents, detailing the history of patient care and doctor intervention, and are often subpoenaed and used as evidence in malpractice cases. Due to this secondary use, we'll mark some things that may not seem relevant to the patient's present illness, but are very relevant to the quality and nature of the patient's treatment and the doctor's handling of the case. As such, we'll mark as timeline-relevant occurrences like discussions with patients, observations by the patient, and any expression of the patient's desires. Thus the verbs "discussed", "compared" and "reports" in the following sentences would all be treated as events in the following examples:

- (11) We <discussed> the risks of this procedure
 (12) I then <compared> her ECG with the one taken on 12-MAY-2001 21:11
 (13) She <reports> severe back pain.

These verbs are not technically relevant to the patient's timeline of treatment. However, the fact that the doctor discussed those risks and compared those ECGs shows, from a legal standpoint, a certain standard of quality, and thus, they need to be included on our timeline. Similarly, recording the fact that the back pain is reported *by the patient* rather than by a nurse or by the doctor helps provide recourse in the event of a legal challenge, and thus, would be considered EVENTS of the type "EVIDENTIAL".

Finally, note that although things like people, places and body parts are obviously not events, there are a few situations when they can be involved in events, such as "no right arm" or "swollen spleen".

Note also that non-clinical states characterizing a person, such as "female" or "coal miner", are not relevant to a timeline and therefore not to be annotated (even if they seem medically relevant), yet age (as in "27 year old man") is very much relevant.

3.2.2 Finding Syntactic Heads

The previous section used <angle brackets> to discuss events that we wanted represented upon the timeline. Yet as noted before, we are not actually annotating entire spans like that. Instead, we will be labeling only one word, and plan to automatically recover the rest of the event from syntax. The following examples, therefore, one might first look at the full events which will show up in the timeline are "coronary artery disease", "diagnosed", and "unable to lift her arms":

- (14) Comorbidities include <coronary artery disease>, <diagnosed> in {July}

(15) She's <unable to lift her arms>

Of these spans, the only parts we label will be the syntactic heads of these events. What you should actually label as EVENTS is therefore shown in [brackets] below:

(16) Comorbidities include coronary artery [disease], [diagnosed] in {July}

(17) She's [unable] to lift her arms

This distinction is between the full span we want represented on the timeline, and the constrained, single-word SYNTACTIC HEAD which we will actually mark. Do not worry if the syntactic head does not capture what you consider to be important about a given event. We will be automatically expanding out from these syntactic heads to capture the full spans of the EVENTS, but for ease of machine discovery of EVENTS, it's simply easier to start from headwords.

As such, it's important to be aware of the rules that follow, in order to really understand what the consequences are of selecting a given syntactic head - what the "head" actually means. We will start with a strict definition of headedness, but numerous examples will also follow throughout the rest of the guidelines. It is hoped that you will have some exposure to the concept of headedness already through linguistics courses (or even grammar school!), but do not despair if you do not.

We start with the traditional notion of headedness: that there is a word in each phrase which essentially defines and represents that phrase. The head is not the same thing as the most important part of a phrase, but is rather the most basic. An initial test for this is that the "head" can generally stand in for the entire phrase, and mean roughly the same thing with roughly the same grammatical properties. Thus in a noun phrase like "my insatiable need for more donuts", the thing that really represents the notion is "need."³

Another way to think about this is to consider what is being modified by the words in this section of the sentence. With "my insatiable need for more donuts", both "my" and "insatiable" give us more information about the nature of the need, and "for more donuts" clarifies more about what the need is for. Chances are, if you look at the whole clause, you'll find that much of the information is pointing to or modifying a single word, and that will usually be your head.

For verb phrases and adjective phrases this will be simple, and there is no need to think of things in terms of SYNTACTIC HEADS at all. You will only be annotating the verb or adjective in the phrase, and nothing more. Thus look at the verbs in the following examples (no other parts have been annotated):

(18) Patient [reports] [throwing] up.

(19) She's [unable] to lift her arms

(20) Since her last [surgery], she has seemed [disoriented] and [moody]

(21) He is [stable], but still [jaundiced]

(22) She feels slightly [weak] but has [resumed] most of her normal activities

(23) Right Meckel's cave is [stable] in appearance with enhancing soft tissue along medial margin.

(24) Urine [Cytologies] were all negative

One may note from the examples above that this approach is taken even when there are modifiers to the verb or adjective, or when one has a multiword predication such as "throw up".

³Depending upon one's linguistic background, some annotators might disagree with that last sentence, as some theories consider determiners like 'the' and 'a' to be the heads of noun phrases or 'DP's. We will not be doing that at all, for practical reasons. One will note that we even ignore measurement phrases like 'some of' or 'three liters of', though the argument for their headedness is less controversial.

Annotating the heads of noun phrases will be the hard part of this task. That being said, although it is important to keep questions of substitutability in mind, English noun phrases are very consistently organized, and therefore the head always falls in the same place. If you find a span of text that seems relevant to the timeline yet nominal, follow the following rules:

- First, remove from your consideration everything that is part of a prepositional phrase. In most cases this should leave you with a single cluster of nouns and adjectives at the left (initial) part of phrase:

~~That recision biopsy analysis of the sigmoid colon from today~~

- Secondly, the remaining elements should cohere as a noun phrase. Take the last/rightmost word within that phrase, ignoring the rest:

~~That recision biopsy analysis of the sigmoid colon from today~~

- That element is the head, and should be selected using the annotation tool and given additional features according to the sections that follow:

That recision biopsy [analysis] of the sigmoid colon from today

There are a number of special cases and additional assumptions to consider when selecting the head, as well. These are:

- For measure phrases like 'three liters of blood', we will ignore the phrase that holds the quantity and instead look at the thing being counted (in this case, [blood]) for the syntactic head. This is an exception to the above rules
- Sometimes in clinical text there are fragmentary phrases which are not exactly in line with normal English syntax, such as "CT chest pelvis". In such cases, try to rephrase the sentence in very normal and formal English, and then tag the head according to that rephrasal. In this case one might say "CT of the chest and pelvis" or "chest and pelvis CT", both of which give you a clear head of [CT].
- One should ignore postmodifying adverbs and adjectives, such as "bilaterally" in "patient had femoral artery grafting bilaterally"
- For coordinated heads, both of the heads will be labeled as separate elements and annotated separately.
- If the rightmost term in a noun phrase (ie the "head") is a particle for a nominalization of a verb, such as 'up' in 'will return for a follow - up', then ignore that particle and annotate the nominalized verb itself.

Such a list of rules gives us the following heads (only showing the nominal heads):

- (25) The [CT] showed a small rectal [abcess].
- (26) The patient reports [nausea] and [vomiting]
- (27) She does note [darkness] of the stools
- (28) She had experienced no [dizziness] until the [start] of [chemotherapy].
- (29) [Levaquin] 750 mg p.o. q. day (will restart today)
- (30) I explained that BRAF [mutations] have no predictive value with regard to cetuximab [sensitivity]
- (31) We also discussed some of the [toxicities] of fluoropyrimidine-based chemotherapy.

It is very important to consider examples such as (31) above. The relevant EVENT that we care about is the phrase "toxicities of fluoropyrimidine-based chemotherapy." As such, even though 'chemotherapy' might look like something you would place upon a timeline, it is contained within the noun phrase marked by "toxicities", *and therefore is not annotated*. This kind of pattern is discussed below. Whenever you have questions about which

parts of a phrase to annotate, pass those questions along – it is important to get agreement and consistency about such decisions.

3.2.3 Some Exceptions and Specific Patterns

Measurement Phrases Finally, there is one exception to those above rules, concerning so called "measure phrases." Although technically units of measure such as "liters" below would be the heads of phrases like "three liters of blood", such measurement phrases are *not* annotated as the syntactic heads for the purposes of THYME. Instead, mark the thing being measured, as with "fluid" below:

(32) He is taking in adequate [nutrition] and adequate [fluids]; consumed 3500 [calories] and drank 2-3 liters of [fluid]

This applies to all such phrases where the quantifier might technically be the syntactic head in normal linguistic analysis, including 'some of' in the following:

(33) We also discussed some of the [toxicities] of fluoropyrimidine-based chemotherapy.

Test Results, Measurements, and Labs When discussing a test and its results, we still maintain our policy of marking the syntactic head.

(34) The CT [scan] was normal

Here, the EVENT is the CT scan itself. The finding of "normal" is captured by comparison with other annotations that link tests and results.

(35) Her [BUN] is 9 mg/dL

Here, the only EVENT is the test itself ([BUN]), and the finding (9 mg/dL) is captured by a later step.

The same goes for formulaic measure sections:

(36) [Height]=178.60 cm,

(37) [Weight]=91.90 kg,

(38) [Systolic]=160 mm/Hg,

Here, we only capture the measure itself, and not the value.

The only exception to not capturing the test result is when the result of a test, lab, or exam is the diagnosis (or suggestion) of another disorder, disease, or disorder, as below:

(39) The CT [scan] [showed] probable [adenocarcinoma]

(40) Her [CBC] [indicates] [thrombocytopenia]

Here, the test is an EVENT, as is the revelation itself, as well as the diagnosis, and all merit inclusion on the timeline. Put differently, a measure isn't a separate EVENT from the test which indicated it, but a diagnosis or disease shown by a test is. Note that the [showed] and [indicates] are both EVENTS in their own right, of TYPE EVIDENTIAL.

Latin Names For Latin names, do your best to mark the Latin head of the phrase. So, for:

- (41) She has [diabetes] mellitus
- (42) [Polymyalgia] Rheumatica is a concern here.
- (43) We suspect [myasthenia] gravis is the cause of her continued weakening.

Generally, the head of a two-word Latin phrase is the leftmost expression, as the second word is often a modifier for the first. If in doubt, mark the leftmost word.

Complex Predications The following kinds of phrases tend to really be a single event, and therefore are *only* headed by the first word when they are noun or adjective phrases.

- Risk of X: *We [discussed] the [stage] of the cancer and that the oncologist {today} would attempt to [quantify] the potential [risk] of cancer recurrence.*
- Inability to X/ unable to X: *Due to the extensive [disease] in the liver, Dr. Solomon was [unable] to proceed with planned hepatic resection and [liver disease] was [left] in.*
- Evaluation for X: *He was undergoing [evaluation] for this and for possible surgery at Newark when he experienced a [stroke].*
- Difficulty doing X: *[Intraoperatively], there were no [difficulties] securing his airway.*

This only applies to nouns, though. Because we do not think of verbs as having a greater span in the sentence, similar phrases using verbs would be given separate EVENT annotations:

He has been taking [narcotics] preoperatively and had [struggled] with [pain] while [hospitalized.]

Separate Predications In contrast, the following kinds of patterns tend to involve multiple events, but these are to be treated as separate events, each given their own EVENT annotations:

Reveals: *[Examination] [shows] decreased bilateral peripheral [pulses].*

Reporting: *and the right hilar [lesions] were not [reported] as being [prominent]*

Denies: *He [denies] any chest [pains], chest [pressures], [shortness]-of-breath, or exertional [symptoms].*

Preference: *Chewing tobacco [cessation] performed. He is [willing] to [meet] with our Pauahi Wing Queens Hospital.*

It's worth noting that EVIDENTIAL EVENTS (such as [shows], [reported], [reveals]) almost always have two separate predications.

Some Tests for Semantically Light Predications can sometimes be hard to identify. Note, as mentioned above, that copular verbs like "to be" or "seems" are not annotated, nor are predications like "experience" or "underwent":

- (44) At that time, he also experienced rectal [pressure] / [fullness] and occasionally had [hematochezia]

Other verbs are not as consistent. For example, we want to treat actual events of needing and lacking as EVENTS, while ignoring other predications which simply add a semantic feature of necessity to the main predication. One (admittedly slippery) test is whether you can rephrase the instance of 'need' using the less eventive 'it is necessary ...' and maintain the same meaning. If so, then it is non-eventive and should not be annotated.

- (45) We may need to [hold] her [diuretic] and/or angiotensin receptor [blocker] if [continued] [improvement] in creatinine is not [noted]. (can be rephrased with "It may be necessary for us to hold her diuretic...")

adeno-	circum-	-ette	inter-	-logist	non-	pseudo-	tele-
aorto-	cis-	ex-	intra-	-logy	novem-	post-	tera-
ante-	colo-	extra-	-ian	-ly	octa-	pre-	tetra-
anti-	contra-	-er	-ible	judeo-	octo-	pro-	tri-
arch-	cortico-	-ery	-ing	macro-	o-kay	quasi-	uber-
ambi-	cran-	ferro-	-isation	micro-	ortho-	quadri-	uh-huh
-able	crypto-	-ful	-ise	mid-	over-	quinque-	uh-oh
-ahol	-centric	-fest	-ising	mini-	paleo-	-rama	ultra-
-aholic	-cracy	-fold	-ism	mono-	para-	re-	un-
-ation	-crat	gastro-	-ist	musculo-	pelvi-	recto-	uni-
axio-	de-	-gate	-itis	neuro-	pheno-	salpingo-	vice-
be-	deca-	-gon	-ization	nitro-	penta-	sero-	veno-
bi-	demi-	giga-	-ize	mm-hm	pica-	semi-	ventriculo-
bio-	dis-	hepta-	-izing	mm-mm	poly-	sept-	-wise
broncho-	-dom	hemi-	ideo-	-most	preter-	soci-	x-ray
co-	eco-	hypo-	idio-	multi-	pan-	sub-	x-rays
counter-	electro-	hexa-	infra-	medi-	peri-	super-	x-rayed
cross-	ennea-	-hood	iso-	milli-	phospho-	supra-	
centi-	-esque	in-	-less	neo-	pneumo-	sur-	

Figure 1: Affixes which can form larger heads

- (46) Blood [labs] revealed an urgent [need] for more iron. (note how hard it would be to rephrase using "it was necessary ...")

Another such test can be used for verbs of planning, which are also of variable semantic weight. If 'plan to' can be replaced with 'will probably' or 'hopes to' without major change in meaning, then it is too semantically light, and should not be tagged. In contrast, if it referring to the clinical act of planning out a patient's future treatment, then it is eventive.

- (47) No oncologic [follow] - up is [planned] at this time.
- (48) Patient is planning to [return] to his local oncologist for which he has a good relationship with. (see rough equivalence with "Patient will probably return to his ...")

Hyphenated Words: A hyphenated word like "CT-scan" will be treated as two separate parts, "CT" and "scan", and will therefore be annotated as if the two parts were separated by a space. As such, "CT-scan" would have a single head, "CT - [scan]". The general rule is that if you could replace the hyphen with a space while still finding the phrase grammatical, then you should pretend that they are separate tokens. If, instead, the prefix or suffix seems inseparably bound to the word, it can be treated as a singular term.

Figure 1 contains a list of affixes which are to be treated as "single words" – one should attempt to stick to this list as much as possible.

Discussions are always eventive, but do not form larger predications - the topics discussed are annotated as separate EVENTS if timeline-relevant. Note that these will be often marked as "generic" in their modality.

- (49) We [discussed] the role for adjunctive [treatment] if this is colon [can- cer] as well as neoadjuvant [treatment] if this is felt more consistent with a rectal [cancer] on flexible [sigmoidoscopy].

Implicit EVENTS Another odd situation which will come up in your annotation are sentences which have implicit EVENTS:

(50) Her main concern is that she does not wish to have a [colonoscopy], which she had back in the 1970s.

Here, there are actually two colonoscopies being discussed. One is a hypothetical, near-future one which the patient is not interested in having, the other occurred in the 1970s. However, implicit EVENTS are not marked in this project, so, you'd have a single annotation with the span [colonoscopy], DocTimeRel FUTURE, Modality HYPOTHETICAL, Polarity NEG.

This same policy of marking only the surface EVENT and ignoring the implicit EVENTS will apply to all of them.

3.2.4 Specific Prohibitions

There are two specific span-types which will *never* be considered EVENTS in our schema:

Anatomical Sites are never annotated as EVENTS, even at cost of leaving certain EVENTS unmarked. So, even though there may seem to be eventive things going on in the following sentences, no EVENT should be marked:

(51) He has a spastic colon.

(52) She complains of sore throat.

(53) She has a missing arm.

All three of these events are unmarked because their syntactic heads (“colon”, “throat” and “arm”) are anatomical structures. Compare these to the below:

(54) His colon is [spastic].

(55) Her throat is [sore].

(56) Her arm is [missing].

Note that this is determined syntactically, rather than semantically, and the only determiner of whether something is marked or not is whether the syntactic head is a piece of anatomy.

Note, though, that tumors and associated cells are not considered anatomy, but pathology, as they are not part of your average human being. So:

(57) Pathology [revealed] some tumor [cells].

(58) John's pancreas is [displaced] by a large [tumor]

(59) Biopsy revealed normal epithelial cells.

Numerical expressions (digits) are similarly never marked as EVENTS. These can be coordinates of tumors, ages, or other numerical expressions, but they are *never* EVENTS:

(60) She was 92 when she had her first [MI].

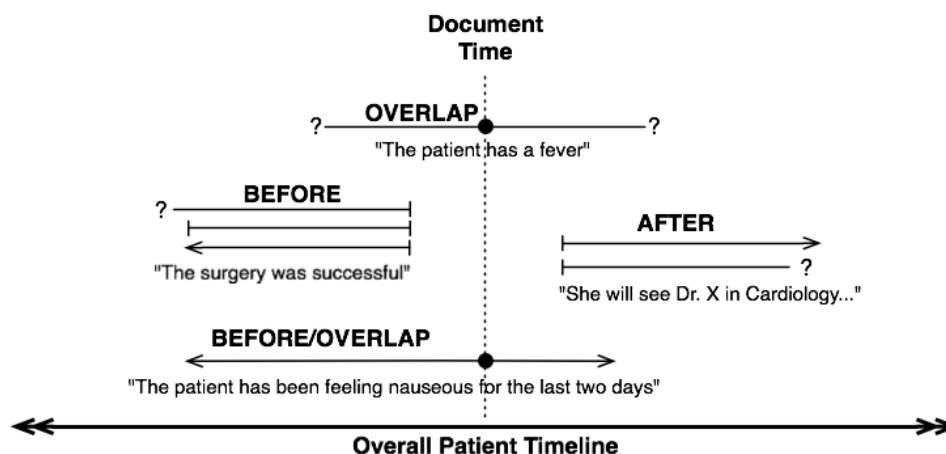


Figure 2: Schematic view of all the DocTimeRel possibilities relative to DOCTIME

- (61) The [mass] is 2.3 cm from the end of the sigmoid colon.
- (62) Her heart [rate] is 62bpm

Note that in the first example, the age is not marked, the location is a property of the mass, and 62bpm is the result of a test captured in [rate].

3.3 Annotating DocTimeRel of EVENTS

DocTimeRel is short for "Document Creation Time Relation", and represents the temporal relation between the EVENT in question and the time when the medical record in question was created (the "document time"). For the purposes of this schema, we are assuming that writing of the record itself is functionally equivalent to the time of the patient's visit to the physician. So, anything considered true during the visit will be considered true when the visit was documented by the physician.

DocTimeRel allows us to avoid the linguistic ambiguities inherent in explicitly marking the grammatical tense of the verb (like "past", "present", or "past perfect"), instead marking the actual temporal relations of the event to the time when the document was created (marked with DOCTIME in this schema). We've chosen to use the same temporal relations present in TLINK annotations.

When annotating DocTimeRel on EVENTS, remember that this is the relation of the EVENT in question relative to the moment in time and space when the record was written. So an event which occurs before the time of writing would be given the BEFORE value, but an EVENT which will occur after DOCTIME will be given the AFTER value for DocTimeRel. Thus we see that marking DocTimeRel on the EVENT can be thought of as a faster, easier way to temporally link the EVENT to DOCTIME (rather than making a TLINK for every event).

Unlike the other EVENT attributes, DocTimeRel has no default value, as it should be filled in individually for every event. Currently, our schema includes four relations between the event and DOCTIME: BEFORE, AFTER, OVERLAP, and the combined relation BEFORE/OVERLAP.

3.3.1 BEFORE

BEFORE is used where the event ended before the patient was seen (and thus, before the document itself was written). The bracketed events below would be marked as "BEFORE" (and all other EVENTS and TIMEX3s are unmarked):

- (63) He is taking in adequate nutrition and adequate fluids; consumed 3500 [calories] and drank 2-3 liters of [fluid]
- (64) This is unchanged and may be related to treatment [changes].
- (65) Today's study demonstrates a marked improvement compared to the prior 9-16-03 [study].
- (66) Until last week, the patient had had no [nausea].
- (67) She had experienced no [dizziness] until the start of chemotherapy.
- (68) The patient had had no [fever] before the start of her [surgery] last week.

3.3.2 OVERLAP

OVERLAP is used for events or states which are happening or true at the time that the patient was seen and thus, we presume, when the document was written:

- (69) The patient [continues] to [do] well as an outpatient.
- (70) The patient is [alert], [cooperative], and appears to be in no [acute distress].
- (71) Moderate sized retention [cyst] or [polyp] in the right maxillary antrum again [noted]
- (72) He is [taking] in adequate nutrition and adequate fluids; consumed 3500 calories and drank 2-3 liters of fluid
- (73) She is not [interested] in pursuing chemotherapy at this time but is interested in continued [surveillance]

3.3.3 AFTER

AFTER is used where the event is scheduled or planned to begin following the document time:

- (74) [Levaquin] 750 mg p.o. q. day (will restart today)
- (75) The patient will [return] tomorrow for [labs] and [exam].
- (76) She is not interested in pursuing chemotherapy at this time but is interested in [continued] surveillance

It's worth pointing out that (76), considered alongside (73), shows the interaction between DocTimeRel and ALINK to cover the idea of "already happening, and will now continue".

There is one specific situation which must be discussed. EVENTS with a Contextual Modality of GENERIC (discussed in 3.7.4) will always have a DocTimeRel of OVERLAP as stated truths are, presumably, true at Document Time. See (77) c

- (77) I explained that BRAF [mutations] have no predictive value with regard to cetuximab [sensitivity]

3.3.4 BEFORE-OVERLAP

BEFORE-OVERLAP is used where an event started BEFORE the DOCTIME) and continues into and through the DOCTIME (OVERLAP with DOCTIME). Simply put, this is used where an event started before the exam or patient

visit and continues through to the present, and often (but not always) corresponds with the use of the English present perfect tense:

- (78) The patient has [felt] quite well and his appetite has been [good].
- (79) She has not [seen] a cardiologist.
- (80) She has had no [fever].
- (81) Based on last Thursday's MRI, her [swelling] is mostly gone.



Caution!

Be sure that any time you're using BEFORE-OVERLAP, it's been explicitly stated that an EVENT started before DOCTIME and continues. Even if you as a human can infer that a given EVENT must've started before the document time, don't mark an EVENT as BEFORE-OVERLAP unless there's evidence in the sentence itself, as the software we're training won't have that benefit.

3.4 Annotating type of EVENTS

Some EVENTS don't actually represent clinical events, but instead, provide aspectual information (starting, stopping, continuing about other EVENTS). To differentiate these EVENTS from the traditional clinical EVENTS which occur on a timeline, we use the "type" marker. It has three values: "N/A", "Aspectual" and "Evidential".

3.4.1 N/A

"N/A" is the default value, and represents the vast majority of EVENTS in the schema, and unless explicitly mentioned otherwise (below or in the ALINK section), all EVENTS used in examples in the Guidelines. Unless the EVENT is of the specific, relatively closed class, listed below under ASPECTUAL, or providing evidential information (see EVIDENTIAL), it will be marked as "N/A" (or, as it is the default value, left blank).

One other note on the word "recurrence", which is often troublesome for annotators working with the N/A/ASPECTUAL distinction. In an example like:

- (82) She has significant risk for tumor [recurrence].

"recurrence" does actually carry some aspectual information (the tumor would have restarted). However, because the word "recurrence" would not be aspectual in "she has risk of recurrence", we have chosen never to mark the work it as an ASPECTUAL EVENT. Instead, this will be an EVENT of type N/A, with a span of "tumor recurrence", as shown above. No ALINK annotations will be made here.

3.4.2 ASPECTUAL

The next EVENT type is ASPECTUAL, which is used to indicate an event whose function is to emphasize or code the aspect of a later event, like "continues" or "restart". Every EVENT of type "aspectual" must later participate in an ALINK.

- (83) The rash has not [reappeared] and we will monitor closely
- (84) His anterior chest rash has not [reoccurred].

- (85) We're going to [hold] her heparin until after her surgery.
- (86) The patient will [continue] treatment.
- (87) She is not interested in pursuing chemotherapy at this time but is interested in [continued] surveillance

These represent a relatively closed class, and you will find yourself marking the same words as aspectual EVENTS over and over again. This is expected, and should not be cause for concern.

3.4.3 EVIDENTIAL

The other EVENT type is EVIDENTIAL. EVIDENTIAL EVENTS are little bit strange, in that they're not technically relevant to the patient, but they provide information about how doctors came to identify and learn about other events. EVIDENTIAL EVENTS are a relatively closed class, generally verbs of showing, demonstration, evidence, confirmation or revelation, and in the clinical domain, are very commonly associated with tests, imaging, and human observation.

In short, an EVENT should be marked EVIDENTIAL only if it serves as the link between a source of knowledge or observation and a piece of knowledge gained from it.

- (88) Her CT-scan [showed] a small mass in the right colon.
- (89) Subsequent bloodwork [revealed] severe anemia, and the patient was admitted
- (90) The patient [reported] severe back pain throughout the month of May
- (91) Dr. Green [pointed] out a patch of skin discoloration suspicious for melanoma
- (92) We will plan proceed with surgery unless her tests [indicate] cardiac problems
- (93) His home health nurse [noticed] an mild increase in confusion and fatigue during chemotherapy
- (94) CT scan [confirms] diverticulitis
- (95) She has a [confirmed] sulfa allergy.

It is worth reiterating that one does not mark the test, reporter, or perceiver as EVIDENTIAL, instead, you'll mark the verb of perception, reporting, revelation, or indication as an EVIDENTIAL EVENT.

Finally, much like aspectual EVENTS, you will be likely marking the same verbs as evidential over and over again, and will very quickly learn to recognize which verbs belong to this class and which do not.

3.5 Annotating polarity of EVENTS

In order to express the polarity of an EVENT, the "polarity" attribute of an event is specified. Polarity in this schema is relatively straightforward, and there are only two possible types: POS and NEG

3.5.1 POS

The first and the most commonly polarity value used is POS. This is the marker of positive polarity. This is used for an EVENT that did, in fact, occur. Most events annotated are of this polarity, and this is the default degree. POS need not be specified in annotation.

- (96) The patient has [hepatosplenomegaly].
- (97) PO [changes] right pterional [craniotomy]
- (98) The patient will [continue] treatment.

3.5.2 NEG

The opposite of POS, as you might guess, is NEG, which is used to indicate when the event didn't take place, or has an otherwise negative polarity:

- (99) No evidence for new suprasellar [mass].
- (100) Otherwise, he has not had any [nausea], [vomiting], [diarrhea], chest [pain], [shortness] of breath, or [fever]
- (101) She is not [interested] in pursuing [chemotherapy] at this time.
- (102) The patient did not [report] [nausea].
- (103) A cystic duct lymph [node] is not [identified]

In (102) and (103), there are two EVENTS being negated. The first is a verbal EVENT of reporting, the second is a condition. In this case, we can negate both the reporting of nausea and the identification of the lymph node, as neither occurred. In addition, the nausea and lymph node aren't present. So, both in both examples are negated.

We should also highlight that NEG means "did not happen" or "not true", rather than "negative" in the medical testing sense (usually meaning "shows no signs of cancer"). So, for something like:

- (104) Her [colonoscopy] was negative.

The colonoscopy is still polarity POS (as it did, in fact, happen), and "negative" is simply telling us that no cancer was found. An actual polarity NEG colonoscopy would be something like:

- (105) We were unable to perform a [colonoscopy] due to bad prep.

Finally, NEG does not imply any sort of atemporality, and negated EVENTS can still be temporally CONTAINED. See Section 11.2.



Caution!

Don't worry about double-negation or phrase-level negation. In the phrase "She denies vomiting or nausea", [denies] is POS (as the denial is real) and [nausea] and [vomiting] are both NEG because they didn't happen. Each EVENT should be considered on its own (rather than as part of a greater denial), and if the EVENT did not happen, it is NEG, no matter what phrasing may have preceded it.

3.6 Annotating degree of EVENTS

In order to express an incomplete degree of an EVENT, the "degree" of an event is specified.

In practice, degree is used as a companion to polarity, as a way of allowing us to say that something is "mostly" or "a little bit" true, rather than forcing us to call every EVENT 100% positive or negative. Effectively allowing something to be "a little true" ("She has slight pain") or "largely false" ("her scar is nearly gone"), allowing greater nuance in our representation of EVENTS than POS or NEG generally allows.

3.6.1 N/A, MOST and LITTLE

Our three different degrees are N/A, MOST and LITTLE. N/A is used where there is no need to mark either of the other two degrees on the EVENT, and is the default value for degree. These are used when there has been "a little" of an event, or a large (but not complete) change:

- (106) There is a small amount of bright T1 [signal]. (LITTLE)
- (107) Abdominal tenderness has nearly [disappeared]. (MOST)
- (108) She feels slightly [weak] but has resumed most of her normal activities (LITTLE)

3.7 Annotating contextual modality of EVENTS

Our current schema has four contextual modalities, ACTUAL, HEDGED, HYPOTHETICAL and GENERIC. Please note that this is unrelated to grammatical mood or modality, and these modalities give information about the modality expressed in the document, not about the grammatical forms used to express them. In some cases, these modalities can represent uncertainty about an EVENT

3.7.1 ACTUAL

The first is ACTUAL, which is used most of the time, and is the default option (that need not be specifically marked). The majority of EVENTS are ACTUAL, having already happened or being scheduled (without hedging) to happen.

- (109) The patient's new [tumor] is 3.5cm from the epiglottis.
- (110) The patient did not [report] [nausea].
- (111) His anterior chest rash has not [reoccurred].

3.7.2 HEDGED

EVENTS are marked as hedged when the doctor mentions a given EVENT with any sort of hedging. This hedging can be lexical ("seems", "likely", "suspicious", "possible", "consistent with"), or phrasal ("I suspect that...", "It would seem likely that"). These EVENTS are strongly implied, but, for safety, liability, or due to lack of comprehensive evidence, are not stated as fact by the doctor. As such, it's very important that these hedged diagnoses and findings be included in the timeline, but be marked so that they can be easily differentiated from hard and fast diagnoses.

- (112) Ultrasound findings were felt to be consistent with a T3, N1 rectal [tumor]
- (113) An approximately 3cm nodular region of intermediate T2 signal involving the body of the corpus callosum is suspicious for [residual or recurrent tumor] but appears unchanged from the patient's prior examination.

- (114) She has a rash not inconsistent with [measles].
- (115) The patient may have undergone a mild [stroke].

Note that a doctor providing or commenting on evidence for a given finding does not qualify as hedging, so the following EVENTS would not be marked HEDGED:

- (116) She [denies]_{ACTUAL} [vomiting]_{ACTUAL, NEG}
- (117) There is no [evidence]_{ACTUAL, NEG} of [MS]_{ACTUAL, NEG}

But further active hedging can push this over into HEDGED:

- (118) She [denies]_{ACTUAL} [bulimia]_{HEDGED, POS}, but all [signs]_{ACTUAL, POS} point to its presence_{HEDGED, POS}
- (119) There is no concrete [diagnosis]_{ACTUAL, NEG} of [MS]_{HEDGED, POS}, but given her [symptoms]_{ACTUAL, POS}, it seems extremely likely.

3.7.3 HYPOTHETICAL

The third modality is HYPOTHETICAL. This is useful when annotating diagnoses, theories, or other medically relevant but hypothetical events. Hypothetical EVENTS will often follow "if" statements ("If X happens, then we'll use Y to treat Z") or other sorts of conditionals ("Depending on the patient's response, we might treat A with B or with C").

- (120) I've warned the patient that this new medication may cause peripheral [edema]
- (121) We suspect either [achalasia] or [pseudoachalasia] here.
- (122) If she experiences a [fever], we'll [treat] it on an outpatient basis.

It's worth noting that an EVENT occurring in the future does not imply that the EVENT is HYPOTHETICAL (although most hypothetical EVENTS will be AFTER DOCTIME). Although it's true that there's always a degree of uncertainty with anything happening in the future, HYPOTHETICAL marks explicit uncertainty in the text, and should not be used just to indicate this future-uncertainty.

For instance:

- (123) We may [recommend] to [resume] the [Cipro] and [Flagyl] and obtain a [CT] of the chest, abdomen and pelvis.

In (123), all the bracketed EVENTS are HYPOTHETICAL and have DocTimeRel of AFTER.

- (124) The patient's [myringotomy] will take place on Friday.

Here, Myringotomy has a DocTimeRel of AFTER, but is ACTUAL. Compare that to:

- (125) If she has additional [bleeding] next week, she should [come] back in

In (125), both "additional bleeding" and "come back in" would be HYPOTHETICAL, in addition to being AFTER the DOCTIME.

In addition, polarity and modality of EVENTS don't interact, even though one might expect them to. It's true, at least from a real-world point of view, that a HYPOTHETICAL EVENT, by definition, hasn't happened. However, polarity shouldn't change on that basis alone. There are POS polarity HYPOTHETICAL EVENTS ("She might develop a rash") as well as NEG polarity ones ("If this medication works, he will have no soreness").

It's also worth noting that very often, verbs of discussion lead to HYPOTHETICAL EVENTS (although they themselves are not hypothetical):

(126) We discussed the risks of adjuvant [chemotherapy] and [hemicolecotomy] with the patient.

(127) Discussed the potential for neoadjuvant [therapy] depending on the findings of his MRI and studies.

But, of course, one must be careful not to overapply the rule, as sometimes actual facts can be discussed as well, as in (128):

(128) I [discussed] the [effects] of his drinking and the [progression] of his liver cirrhosis.

Note that in (128), the [effects] are ACTUAL, but the [progression], at this point, is hypothetical and contingent on the patient's continued drinking. Please see Section 11.3 for additional detail about the usage of HYPOTHETICAL with verbs of discussion.



Caution!

Although many human languages feature such an interaction, in our schema, modality does not interact with polarity or DocTimeRel. Future EVENTS are assumed to be actual unless stated otherwise, and it's possible (even common) to have an ACTUAL negated EVENT. Future or negated do not automatically mean "hypothetical".

3.7.4 GENERIC

GENERIC is our fourth contextual modality, and is used for EVENTS which may be mentioned in a note, but are only mentioned in a general sense, and do not appear on the patient's timeline of treatment. These usually occur when the doctor is putting justifications of decisions, rationales for changing care, or simply covering his or her back, and often you'll get several sentences of "general discussion" which could have as easily come from a textbook as a medical record. Please see Section 10.1 for additional detail about the usage of GENERIC in notes.

All of the below EVENTS would be marked GENERIC under our schema:

(129) Adjuvant [chemotherapy] following [surgery] is generally recommended in situations similar to this.

(130) I explained that BRAF [mutations] have no predictive value with regard to cetuximab [sensitivity]

(131) In other patients without significant [comorbidity] that can [tolerate] adjuvant [chemotherapy], there is a [benefit] to systemic adjuvant chemotherapy."

Although HYPOTHETICAL and GENERIC may seem similar, remember that most HYPOTHETICAL EVENTS still refer to the specific patient's care, but depend on some eventuality occurring, whereas GENERIC EVENTS could appear in any patient's note, and could just as easily appear in a journal paper.

Finally, remember that as discussed in 3.3.2, **if an EVENT is GENERIC, DocTimeRel will always be OVERLAP.**

3.8 *Annotating contextual aspect of EVENTS*

Aspect is used to express aspectual ideas about the events which are not coded explicitly with aspectual EVENTS and ALINKs. We have three values for contextual aspect in the schema, N/A, INTERMITTENT and NOVEL. Please note that this is unrelated to grammatical aspect, and these two aspects give information about the temporal relations in the document, not about the grammatical forms used to express them.

3.8.1 N/A

N/A is the most common value (and our default value) for contextual aspect, and simply represents that a given EVENT is neither NOVEL nor INTERMITTENT. If neither of the other contextual aspects seems to fit for a given EVENT, leave the contextual aspect field blank (which will then be auto-filled with N/A).

3.8.2 NOVEL

The NOVEL indicates, well, novelty, and is associated with predicate adjectives like "new".

- (132) The patient's new [tumor] is 3.5cm from the epiglottis.
- (133) The newest [MRI] revealed a previously undiscovered mass.
- (134) She's experienced unusual soreness around her new [stitches]

3.8.3 INTERMITTENT

INTERMITTENT is used in situations where there may be a series of smaller events within a single EVENT, rather than a single, constant event. Those events are usually marked with words like "intermittently" or "occasionally", and when such phrasing is used, the EVENT is marked as INTERMITTENT. These indicate that, for instance, the patient has had vomiting since a certain time, but he/she has not been vomiting 24/7 since that point.

Please note that INTERMITTENT will only be used for irregular, unpredictable periods (like the span between seizures or vomiting) and not for things like medications or dialysis (which occur on a set schedule). An event which occurs at an explicit interval can either be treated as a constant ("the patient is taking montelukast for asthma") or the interval can be marked as a TIMEX3 ("She undergoes dialysis [every three days]"), which is then TLINKed to the original event.

It's important to note that we are only marking INTERMITTENT when there is an explicit mention of intermittency in the sentence. Even if you happen to know that a given disorder or symptom often manifests intermittently, if it's not stated explicitly as doing so, you should not mark it as such. As with all of these annotations, we are marking the relations mentioned in the document, not those you can infer from your own background knowledge.

- (135) He reports occasional bright red [bleeding] from the rectum.
- (136) Patient complains of intermittent chest [pain]
- (137) She's had intense [headaches] on and off since her last visit.

If you are unsure about the contextual aspect of a given EVENT, mark it as N/A.

3.9 Annotating permanence of EVENTS

Because we'll be creating a timeline from this data, sometimes, it's valuable to know whether a given condition or disease is FINITE (it might go away with treatment), or is PERMANENT (and the patient will have it for the rest of their lives).

3.9.1 FINITE, PERMANENT and UNDETERMINED

This is a property of the event and disease itself, not of the patient's specific disease. This means that diseases like pneumonia or different cancers are always treated as FINITE even if the individual patient is likely to die before the disease can be resolved. This is as opposed to a condition like Multiple Sclerosis or Diabetes, which is unlikely to be completely cured (given the state of medicine at the time of writing) and is thus PERMANENT. A disease that the patient will have for the rest of their lives, despite any medical intervention, is classified as PERMANENT.

This is medical-domain specific knowledge, and as such, the default value is UNDETERMINED. **Unless you've been specifically instructed to do so, you should not mark permanence for EVENTS.**

4 TIMEX3 Annotation

Because we're looking specifically at temporal relations, the next step of the annotation process is to find and annotate TIMEX3 objects. These are definitive references to time, and will provide concrete temporal references throughout the document or section. Examples of these might be phrases like "today", "tomorrow", "24 hours ago", "at this time" and "early March". In addition, specific dates are annotated as TIMEX3 objects as well.

4.1 Identifying and Annotating TIMEX3s

Our approach to marking TIMEX3s is identical to that used in ISO-TimeML, and these guidelines are heavily based on the standard established in [1].

Unlike with EVENTS, we will not be selecting headwords only. Instead, syntactically speaking, all TIMEX3 annotations should correspond to a:

- Noun phrase ("this weekend", "tomorrow", "yesterday", "Tuesday", "Last May", "May 16th", "6/9/1985").
- Adjective phrase ("two-hour-long", "half-hour" – as in "a half-hour trip", "preoperative", "post-partum")
- Adverbial phrase ("lately", "recently", "shortly", "hourly", "intraoperatively").

Importantly, this means that any prepositions which precede (or in some cases, follow) a temporal expression are to be left unmarked, *even when they seem to provide additional context for interpreting the TIMEX3*. For example:

(138) During {the month of July}, she will come visit.

(139) From {May 1st} to {the 3rd}, she will refrain from eating solid food.

(140) After {tomorrow}, she should be OK to walk with crutches.

(141) He'll come in for a follow up {two days} after his surgery, and again {the next week}.

These prepositions (referred to as SIGNALs in ISO-TimeML), although important in the interpretation of the meaning of the temporal expressions, provide separate temporal information which will be automatically extracted later.

Note that post-expression adverbs (often “ago”) are still captured in our spans, so:

(142) Patient s/p lumpectomy {2 yrs ago}

For the most part, if you have two separate temporal expressions, they’ll be two separate TIMEX3s, but two adjacent TIMEX3s which together specify a single time can be treated as a single span. Look at the contrast in (143):

(143) a. I’ll come by to check on her at {3:30pm Tuesday}
 b. I’ll come by to check on her at {3:30pm} and on {Tuesday}

In (b), we know that the doctor is referring to two different timepoints, so we mark two TIMEX3s, whereas in (a), the “3:30pm” and “Tuesday” combine to specify a single timepoint.

The only exception to this rule is when the two temporally connected TIMEX3s are syntactically separated, as below:

(144) On {Tuesday}, I’ll come by to check on her at {3:30pm}.

(145) She’ll come in for another consult {the day} after {tomorrow}.

In these cases, mark two separate TIMEX3s, even though they combine to specify one timepoint. The sole exception to this rule comes with long-form times:

(146) Mr. Mullins arrived at {5 minutes to five}.

But in medical texts, these are extremely rare.

Finally, as one might imagine, TIMEX3s which are separated by a conjunction are to be treated separately as well:

(147) She should be fine for discharge on {Tuesday} or {Wednesday}.

(148) He will come in on {the 1st} and {the 5th} for followups.

4.2 Annotating TIMEX3 class

Note that in the examples below, if there are multiple TIMEX3s, only those which are of the indicated class will be marked.

4.2.1 DATE

The majority of TIMEX3 annotations you make will be of the class DATE. DATE represents dates. These can be calendar dates (such as {January 4}) and other verbal expressions which can be mapped to calendar dates either concretely (such as {Last week}, {This month}, {next Friday}, or {this time}), or in a more fuzzy sense ({lately}, {the past}).

(149) MRI of the brain without and with gadolinium contrast utilizing tumor followup protocol compared with prior studies of {29, February 2005} and {28, January 2005}.

(150) His anterior chest rash has not reoccurred since the PCN VK was discontinued {24-hours ago}.

- (151) At {this time}, we see no reason to discontinue the treatment.
- (152) The last cyclosporine level was 373 in {January}. His dose was adjusted downward to 300-mg twice-daily. A cyclosporine level will be repeated on {Friday morning}.
- (153) I stated that if there is no other evidence of any disease recurrence, in {approximately one-month's time} we would proceed with approximately six-months worth of adjuvant therapy.
- (154) The form was signed and scanned on {December 29, 2009}.
- (155) I would probably restart her furosemide {tomorrow morning}.
- (156) She came in {a couple months ago}
- (157) She had an appendectomy {two-to-three years ago}
- (158) Mr. Zegler was seen in the Hamilton University Medical Center with Dr. Carr {the middle of December 2009}
- (159) Carotid artery disease, last checked {greater than two years} prior.
- (160) After {next week}, we'll see where her pain level is.
- (161) {June 6, 2008}, through {October 6, 2008}, treated with FOLFOX

As mentioned above, DATE is also used for very generic sorts of TIMEX3s, where you may not be able to point at a specific day, week, or month on a calendar, but can still gesture at the overall timeline. So, for instance, expressions like "in the past", "lately", "in the future", or even "previously", would be TIMEX3s of the type DATE:

- (162) She has experienced heavy bleeding in {the past}.
- (163) She complains that she's felt tired {lately}.
- (164) {Recently}, she has had several episodes of syncope.
- (165) He wasn't sure he wanted surgery at {that point}

4.2.2 TIME

TIME is used for specific time points within a day, for instance, {3:00PM} or {23:45}, and once again can be relative (as in example (167)):

- (166) The patient's MRI is scheduled for {5:30pm}
- (167) Following the patient's latest seizure, {20 minutes ago}, we are re-evaluating her medications.
- (168) Patient did well {overnight}.
- (169) At {the time of consultation} there is no operative report or pathology available.
- (170) Surgery will need to be completed by {2:45} to have the biopsy to the lab sooner.

Put differently, temporal expressions which give minute-by-minute or hour-by-hour detail are marked as TIME. Day-by-day (or larger) detail is marked with DATE.

4.2.3 DURATION

Sometimes, you'll be given a single temporal expression interpreted as reflecting a span of time, rather than a point. These are things like "for {24 hours}" or "All of February", and these are marked with the class DURATION.

- (171) The patient continuously experienced nausea for {nearly two weeks}.
- (172) For {the next 12 hours}, we will lower the patient's morphine drip and then we will re-evaluate his pain.
- (173) Since {August}, she has not had any episodes.
- (174) During {the last 12 months}, she's been nauseous frequently
- (175) Since {February 18th}, he hasn't seen his doctor.
- (176) In {the last week}, his pain has significantly worsened.
- (177) He has been doing this for {five years}.
- (178) I stated that if there is no other evidence of any disease recurrence, in approximately one-month's time would proceed with {approximately six-months} worth of adjuvant therapy.
- (179) In {the time} between now and the 15th, she should attend physical therapy whenever possible.

Note that in (179), both {now} and {the 15th} would also be TIMEX3s, but of type DATE.

Remember again that more abstract temporal expressions ("lately", "in the past", "in the future"), although representing loosely defined spans of time, are considered DATE rather than DURATION, as they are predominantly only bounded at the start or the end, not both, as above.

Finally, remember that two dates can be used to construct a duration, but, because each represents a single point in time (rather than duration), both will still be labeled DATE, rather than DURATION:

- (180) From {May 1st}_{DATE} to {the 3rd}_{DATE}, she will refrain from eating solid food.

4.2.4 QUANTIFIER

Although it may seem odd at first, expressions like "Twice", "four times", and "18 times in the month of May" are all TIMEX3s. These are annotated with the QUANTIFIER class.

- (181) The patient vomited {twice} before the surgery.
- (182) We have seen Mr. Lastname {three times} for his ulcerative colitis.
- (183) On {two to three incidents} she has had blood in the stools

Note that Quantifier only applies for number of occurrences of an EVENT, not for quantifiers like "She has *two* eyes" or "She [lost] 5 units of blood".

4.2.5 PREPOSTEXP

Similarly odd, Pre- and Post- expressions ("preoperative", "post-exposure", "post-surgery", "prenatal", "pre-prandial") all actually designate specific temporal spans ("The time before the surgery", "The time after exposure") related to an implicit EVENT, and thus, are TIMEX3s, marked with the class PREPOSTEXP. Usage of this TIMEX3 is discussed in more detail in Section 11.5.

(184) Patient underwent a partial hemicolectomy in July 2009. {Postoperative} scarring noted during exam.

(185) The patient exhibits {post-exposure} changes.

These will not always begin with "pre-" or "post-". Terms like "intraoperatively" can sometimes be PREPOSTEXP as well, as in (186):

(186) {Intraoperatively}, there were no difficulties securing his airway. He received 4 liters of crystalloid, made 300 mL of urine and estimated blood loss was 50 mL. He received a dose of DDAVP in the OR out of concern for uremic platelet dysfunction

And sometimes, you'll have bare expressions which clearly express the PREPOSTEXP meaning, but don't contain the whole expression.

(187) Pulmonary recommendations for {pre}, {peri} and {postop} were made.

4.2.6 SET

SET is used exclusively in our schema for covering expressions which give both a quantifier and an interval (like "Three times weekly", "monthly" or "1/day") and represent a frequency. This is different from QUANTIFIER ("twice") which only gives a quantifier, and different from DURATION ("all week") which only gives a span.

Even those most sets could be interpreted as a QUANTIFIER and a DATE/TIME juxtaposed, but we should mark them as a single span ("{twice monthly}" rather than "{twice} {monthly}").

(188) Will administer Lariam {twice daily}.

(189) Patient has checked into the ER {roughly three times a month}.

(190) We will proceed with {weekly} consultations to monitor the patient's condition.

(191) Mirtazapine REMERON 7.5-mg tablet 1 tablet by mouth {every bedtime}.

(192) Simvastatin ZOCOR 20-mg tablet 1 tablet by mouth {one-time daily}.

TIMEX3s of type SET should always be TLINKed to EVENTS using the TLINKs of the type OVERLAP

Part II

Relations Annotation Stage

The next step of THYME annotation is to take those EVENTS and TIMEX3s annotated in EVENT annotation and mark the temporal relations between them. By "temporal relations", we're referring to a relatively limited set of

timeline relations between EVENTS and TIMEX3s, referred to as TLINKs ("Temporal Links"). These relations exist for one reason: to allow us to arrange and order the various EVENTS and TIMEX3s accurately on a timeline.

Although these temporal links are annotated in a more interactive fashion, we will display them in this document using the following format:

(193) [EVENT1] RELATION [EVENT2]

Where RELATION is BEFORE, CONTAINS, OVERLAP, BEGINS-ON, or ENDS-ON, as determined by the type of relation being stated. These types are described in detail in Section 6.1.

To give a realistic example, imagine the following sentence:

(194) The patient was to follow-up with oncology this month

The EVENT [follow-up] is clearly related to the TIMEX3 {this month}, because [follow-up] will occur during {this month}, giving you more specificity than the followup just occurring after document time. So here, you would create a TLINK annotation, insert [follow-up] into the "target" slot, {this month} into the "source" slot, and select CONTAINS as the relation. By doing so, {this month} is established as a narrative container anchor, which may contain additional EVENTS later on in the note. So, the finished annotation would look like:

(195) The patient was to follow-up with oncology this month
a. {this month} CONTAINS [follow-up]

These temporal links (along with the rarer "aspectual link" discussed in Section 7) are used to annotate the document, and eventually, to reconstruct the patient's timeline.

But before we discuss the specifics of these temporal links, we need to discuss our strategy for annotation, and a crucial metaphor which, once understood, will yield a higher quality of annotation, while saving us a great deal of time and effort.

5 Narrative Containers

One very important concept for the marking of temporal relations is that of the narrative container, discussed extensively in Pustejovsky and Stubbs 2011 [2]. A narrative container can be thought of as a temporal bucket into which an EVENT or series of EVENTS may fall. These narrative containers are often represented (or "anchored") by dates or other temporal expressions (within which a variety of different EVENTS occur), although they can also be more abstract concepts ("recovery" which might involve a variety of EVENTS) or even EVENTS themselves (many other EVENTS can occur during a surgery). In short, rather than marking every possible TLINK between each EVENT, we instead try to link all EVENTS to a narrative container, and then link those containers so that the contained EVENTS can be linked by inference.

It's worth explicitly stating, though, that these narrative containers are not anything which need to be marked explicitly by annotators, as they're a property of the timeline itself. Instead, the idea of these containers are being presented as an aid to visualizing the temporal relations within a document. If the annotator does his or her job correctly, properly implementing DocTimeRel and creating proper TLINKs, a good understanding of the narrative containers present in a note will naturally emerge from the annotated text.

5.1 Why Narrative Containers?

Our reasoning for using narrative containers is simple: Temporal relations annotation is prone to never-ending webs of temporal links between different EVENTS, because, simply put, every EVENT that ever happened in the

history of Earth is part of a valid temporal relation with every other.

So, for the (somewhat unlikely) paragraph, we could make the following temporal links:

- (196) Napoleon was exiled to Elba in 1814. This has little to do with Mr. Chen's rash and subsequent leg pain, which both developed after his surgery in 2009.
- a. {1814} CONTAINS [exiled]
 - b. {2009} CONTAINS [surgery]
 - c. {2009} CONTAINS [rash]
 - d. {2009} CONTAINS [pain]
 - e. {1814} BEFORE [surgery]
 - f. {1814} BEFORE [rash]
 - g. {1814} BEFORE [pain]
 - h. [exiled] BEFORE [pain]
 - i. [exiled] BEFORE [rash]
 - j. [exiled] BEFORE [surgery]
 - k. [exiled] BEFORE {2009}
 - l. [surgery] BEFORE [rash]
 - m. [surgery] BEFORE [pain]
 - n. [rash] BEFORE [pain]

As we can see, not only is the temporal linking of Napoleon's exile to each modern EVENT absurd, but a number of the TLINKs described above are accurate, but unsatisfying, simply existing "to complete the picture".

Using narrative containers, we can instead put these items into two temporal buckets, {1814} and {2009}, and then *infer from the known relation between the containers* the relations between the items which reside within them:

- (197) Napoleon was exiled to Elba in 1814. This has little to do with Mr. Chen's rash and subsequent leg pain, which both developed after his surgery in 2009.
- a. {1814} CONTAINS [exiled]
 - b. {2009} CONTAINS [surgery]
 - c. {2009} CONTAINS [rash]
 - d. {2009} CONTAINS [pain]
 - e. [surgery] BEFORE [rash]
 - f. [surgery] BEFORE [pain]
 - g. [rash] BEFORE [pain]

In doing so, we can cut the number of required TLINKs in half, while still capturing all of the information of the fully-expanded version using straightforward temporal inference.

To give more concrete examples of this somewhat abstract concept, let's look at some of these containers in the wild.

5.2 Narrative Containers from DocTimeRel

All documents have at least three narrative containers, marked by DocTimeRel. One bucket contains all EVENTS which occur BEFORE the document time, one contains all EVENTS which occur AFTER the document time, and one contains all EVENTS which overlap the document time. Take, for example, this very simplified note section:

- (198) The patient recovered well after her first surgery to remove the adenocarcinoma, and today we reviewed the pathology results and the results of her follow-up CT. Based on these results, pending her consult with Dr. Hart, I feel she is fit to undergo her second surgery in three weeks.

<u>BEFORE DocTime</u> recovered follow-up CT	<u>Dec 16th</u> surgery remove	<u>Jan 3rd</u> admitted fever treated	<u>OVERLAP DocTime</u> reviewed	<u>In 3 Weeks</u> 2nd surgery	<u>Apr. 19th</u> 3rd surgery resection biopsies
--	--------------------------------------	--	------------------------------------	----------------------------------	--

Table 1: Narrative containers and EVENTS in (199)

Here, the red-underlined portion of the note all occurs in a narrative container which ends at the day of the visit, the blue portion all occurs during the course of the visit, and the green section all occurs after the visit, in a narrative container which is bounded by the visit. In this case, all of this information is captured explicitly through DocTimeRel, and the boundaries of these narrative containers are all set by the boundaries of the patient's visit which is being described in the present note.

Most importantly, because we have these separate containers established by DocTimeRel, *we need not create temporal links between these items*, as those links can all be easily inferred by our knowledge of the ordering of the containers. This is the basis of the first TLINKing rule in Section 6.2.

5.3 Annotating additional narrative containers

However, there can be many more narrative containers in a given note than just "Before DOCTIME", "Overlapping DOCTIME", and "After DOCTIME". Take, for example, this note:

(199) The patient recovered well after her first surgery on December 16th to remove the adenocarcinoma, although on the evening of January 3rd she was admitted with a fever and treated with antibiotics. Today we reviewed the pathology results and the results of her follow-up CT. She will have her second surgery in three weeks, followed by a third surgery on April 19th which will involve both additional resection as well as the taking of additional biopsies.

In example (199), there are actually six containers, the contents of each marked again with differently colored underlines. The first surgery occurred entirely within the December 16th container (the anchor TIMEX3), the fever, admission and antibiotics occurred on January 3rd (and within that container), the followup CT and her recovery occurred at an ambiguous point between the surgery and the DOCTIME, the during-visit container, the second surgery occurs three weeks from the DOCTIME, and the final container involves everything occurring on April 19th (the surgery, resection, and taking of biopsies). This clarification of narrative containers, showing which EVENTS happened in which containers, as well as the temporal relations between them, is the role of TLINKs.

In this, more complex example, DocTimeRel alone is not sufficient to mark the temporal relations between all of the EVENTS and to clearly associated all of the EVENTS the narrative containers which they are a part of. Additional detail is needed here to capture that in addition to being BEFORE the Document Time, the first surgery happened on December 16th, whereas the fever occurred on January 3rd. We could also state that the follow-up CT occurred after December 16th. We also need to know that, for instance, the April 19th surgery will happen AFTER the surgery in three weeks. These containers and their events are shown in list format in Table 1.

Again, the critical function of narrative containers is that using them conscientiously allows us to infer TLINKs, rather than forcing annotators to explicitly mark them. Here, there's no need at all to explicitly mark that, for instance, [fever] is before [resection], because we know that the fever happened on Dec 16th, the resection happened on April 19th, and that December 16th is before April 19th. With explicit grouping of EVENTS in this way, we can easily infer the links between them with minimal annotator effort.

As you annotate, it's critical to keep this notion of narrative containers in mind, as to avoid wasting effort marking temporal relations which are already captured straightforwardly.

<u>During Surgery</u>	<u>During CT</u>	<u>During Followup</u>
adhesions	bleeding	CBC
scarring		remove

Table 2: Narrative containers and EVENTS in (201)

5.4 EVENTS as Narrative Container Anchors

So far, we've been discussing temporal expressions (TIMEX3s) as providing the boundaries of narrative containers, but EVENTS themselves can provide boundaries too. Take this example:

(200) During the surgery, the patient experienced another MI, and repeated bouts of tachycardia

In example (200) we can see that the surgery itself is a container, containing an MI (myocardial infarction or "heart attack") as well as tachycardia. In this case, for the temporal structure of the note to really be captured, both the MI and Tachycardia need to be linked to the surgery, using TLINK annotations.

Worth noting is that EVENTS do not have inherent temporal relations with one another in the same way that dates and times do, and must be related for their narrative containers to provide effective inference. Take, for example, the below paragraph:

(201) During her surgery, Mrs. Hayes showed some adhesions and scarring in the abdominal cavity. She did show some bleeding at the incision during her subsequent CT scan. For her followup, we'll take a CBC and remove her stitches.

Here, we have three narrative containers, the surgery, the CT scan, and the followup, with no TIMEX3s at all, all shown in Table 2, above. To fully annotate this passage, in addition to marking that the EVENTS are contained within the surgery, CT, or followup, we'll also need to explicitly mark the relation between those three EVENTS. So, in addition to the containment links (e.g. [surgery] CONTAINS [adhesions]), we'll need to create two TLINKS to relate the containers, [surgery] BEFORE [CT], and [CT] BEFORE [followup].

Once this is completed, inference can easily take place across the containers, and the full benefits of our narrative container approach can be reaped.

5.5 Single-bounded narrative containers

Although so far we've been discussing the idea of containment (and thus, the CONTAINS TLINK) when discussing narrative containers, one can still imagine and annotate narrative containers which are bounded only on one side. Take, for instance, the below example:

(202) Immediately after her [surgery], the patient began to [vomit] and developed a high [fever]

In this case, the [surgery] forms one temporal boundary for the vomiting and fever, and as such, can be thought of as anchoring a one-sided narrative container. In this case, we would say that both [vomit] and [high fever] have a BEGINS_ON TLINK relation to the [surgery]. Although somewhat conceptually different, narrative container relationships can very easily stem from BEGINS_ON, ENDS_ON, and even BEFORE, and the metaphor of the narrative container is still useful in annotating these situations.

5.6 Nested Narrative Containers

Sometimes, you'll find yourself with a situation where a narrative container seems to, itself, belong inside another narrative container. Here's an example with three levels of nesting:

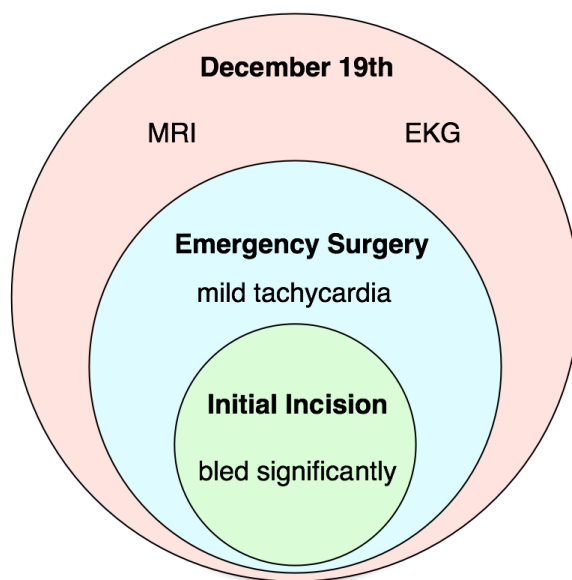


Figure 3: Schematic view of Example (204) as a Venn diagram

- (203) December 19th: The patient underwent an MRI and EKG as well as emergency surgery. During the procedure, the patient experienced mild tachycardia, and she also bled significantly during the initial incision.

Here, we can see that in addition to the overall container of December 19th (containing surgery, an EKG and an MRI), the surgery itself is a container, containing some mild tachycardia and an initial incision. The incision itself forms a third narrative container, containing the significant bleeding.

To annotate a case like this, we'll think about nesting anything else. Whether you're playing with Russian nesting dolls, packing boxes-within-boxes, or packing a car for a long trip, smaller items go in smaller containers, which go in bigger containers, which go in bigger containers, and so forth. Here, we'll use the same principle, but with narrative containers instead of physical objects.

The largest temporal span (actual clock duration) here is "December 19th", so that should be the ultimate narrative container for all these EVENTS. First, we'll put everything into it (surgery, MRI) that belongs there using three TLINKs. The surgery is also a narrative container, but temporally, has a smaller temporal span than December 19th (likely a few hours rather than a full day). Inside it, we'll place mild tachycardia and initial incision using two more TLINKs. Then, finally, the initial incision, with the smallest duration of all, CONTAINS bled significantly. Example (204) shows all the TLINKs required to fully annotate this sentence, representing [EVENTs] in brackets and {TIMEX3s} in curly braces:

- (204) {December 19th}: The patient underwent an [MRI] and [EKG] as well as emergency [surgery]. During the [procedure], the patient experienced mild [tachycardia], and she also [bled] significantly during the initial [incision].
- a. {December 19th} CONTAINS [MRI]
 - b. {December 19th} CONTAINS [EKG]
 - c. {December 19th} CONTAINS [surgery]
 - (i) [surgery] CONTAINS [tachycardia]
 - (ii) [surgery] CONTAINS [incision]
 - [incision] CONTAINS [bled]

Because of the nesting, only the TLINKs mentioned above are necessary. One does not need to, for instance, explicitly link "bled significantly" to December 18th, because it is clearly contained within it by virtue of its container being contained within it. We can visualize these nested containers with a Venn diagram, as in showing these containment relationships, and thus, why we don't need to link the contents of each container to every other container.

Another excellent example of complex narrative container nesting, fully annotated, can be found in Example (233).

So, even in this complex case, we can completely capture the temporal relations present by first putting all EVENTS into their proper containers using TLINKs, and then linking the containers themselves, TLINKing the temporally smallest container to the next smallest, and so forth, until all containers are contained.

This kind of nesting is the only way that a EVENT will ever CONTAIN a TIMEX3, and follows once again our rule of nesting smaller temporal spans inside larger ones. See example (205) below, showing both an example sentence and all the TLINKs needed to annotate it:

- (205) {December 28th}: The patient experienced a [stroke] at {approximately 9:30am}, during her [surgery].
- a. [stroke] OVERLAP {approximately 9:30am}
 - b. {December 28th} CONTAINS [surgery]
 - (i) [surgery] CONTAINS [stroke]
 - (ii) [surgery] CONTAINS {approximately 9:30am}

Here, we have an overarching container ({December 28th}) which CONTAINS [surgery]. The surgery then CONTAINS both [stroke] and {approximately 9:30am}. Then, to complete the annotation, we'd mark that [stroke] OVERLAPS {approximately 9:30am}.

5.7 Choosing the Anchors of Narrative Containers

When creating narrative containers as discussed above, you'll need to choose either a TIMEX3 or an EVENT to be the "anchor" of the container, the temporal span which all of the other EVENTS fall within (or begin/end on). Choosing which temporal span to consider the anchor of a given narrative container to be can be difficult, so here are a few ground rules to help make these decisions easier and more consistent:

1. **The majority of TIMEX3 annotations will be narrative container anchors**
 - Usually, when a date (or other temporal expression) is given in a note, it's to provide temporal context for one or more other EVENTS. As such, they're ready candidates for narrative container anchor.
2. **If you have a choice between using an EVENT or a TIMEX3 as the narrative container anchor, you should pick the TIMEX3.**
 - This is not to say that you'll never have an EVENT within an EVENT, but often that will be because there's no TIMEX3 or because there's narrative container nesting going on, as in Example (204).
3. **If you use an EVENT as a narrative container anchor, try to TLINK it to a few other container anchors to avoid it being stranded**
 - See Section 5.4 for more detail here.
4. **All other things being equal, an EVENT or TIMEX3 with a larger temporal span is more likely to be a narrative container anchor**
 - This is per Section 5.6, where this principle is shown with detailed examples ((204), (205)). This is also visible in example (230).
5. **EVENTs will very seldom CONTAIN TIMEX3s.**

- Only in minute-by-minute descriptions will an EVENT serve as a narrative container in which a TIMEX3 will be included, and even still, there will likely be an overarching TIMEX3 which CONTAINS the EVENT. See example (205) above as well as the discussion in 5.6.

6. Sub-events will be anchored to their main event, assuming there is actually temporal containment

- If the note describes three different steps (e.g. incision, tumor removal and closing) within a single surgery, those steps will be CONTAINED by the surgery. This is the case in all situations where we have an event/sub-event relation (as between "incision" and "surgery" in (204), or "colonoscopy" and "removal of polyps" in (230))

5.8 Ordering within narrative containers

Because of the difficulty of capturing detail within a given narrative container, not all relations between EVENTS will be captured. For instance, in example (204), we do not explicitly give the relationships between EKG and Surgery, nor between the MI and Tachycardia. In many cases, those relationships within a narrative container will follow a set progression (first colonoscopy, then biopsy (which occurs during the colonoscopy), then pathological analysis, then histology), but you as an annotator are not asked (or allowed) to make explicit those domain-specific progressions.

So, although it may seem like some of the narrative containers that you define may be a loose bucket of temporal EVENTS, that's not actually a problem, as the greater timeline of the patient's care is more important than the fine structure within a given narrative container, and as you'll find out, there are still plenty of TLINK annotations to be made.

Of course, as described in 6.2, **if a relation is explicitly stated, it should always be marked.**

6 TLINK Annotation

TLINKs, as previously mentioned, are relations you can mark between EVENTS, between TIMEX3s, or across the two categories to show the temporal relations present within the document and to clearly define the bounds of the narrative containers at work beyond what DocTimeRel will naturally give us. As mentioned previously, we will display these links in this document using the following format:

(206) [EVENT1] RELATION [EVENT2]

Where RELATION is BEFORE, CONTAINS, OVERLAP, BEGINS-ON, or ENDS-ON, as determined by the type of relation being stated. Before we talk about exactly when to use these TLINKs, we'll first discuss the different types of TLINKs used in this schema.

6.1 TLINK sub-types

There are five different temporal relations often used in this schema, BEFORE, OVERLAP, BEGINS-ON, ENDS-ON and CONTAINS. There is no default relation type for TLINKs

6.1.1 BEFORE

BEFORE is fairly straightforward, and they simply order two events in time.

(207) She [vomited] shortly before [surgery]
a. [vomited] BEFORE [surgery]

- (208) His anterior chest [rash] has not [reoccurred] since the [PCN] VK was [discontinued] {24-hours ago}.
- [discontinued] BEFORE [reoccurred]_{NEG}
 - [discontinued] OVERLAP {24-hours ago}

It's worth noting that in (208), [VK] and [rash] are linked to the rest of the EVENTS by using ALINKs, described in Section 7

When annotating, remember that "X occurred after Y" can be expressed by saying "Y occurred before X":

- (209) She was [seen] by Dr. Jones in cardiology following the stent [placement]
- [placement] BEFORE [seen]
- (210) He had a car [accident] shortly after his [visit]
- [visit] BEFORE [accident]
- (211) He had a [neckache] after [surgery]
- [surgery] BEFORE [neckache]
- (212) He had [swelling] after [surgery]
- [surgery] BEFORE [swelling]

6.1.2 CONTAINS

CONTAINS signals that the EVENT is completely contained within the span of the EVENT or TIMEX3 it's related to. In other words, the contained event occurs entirely within the temporal bounds of the event it's contained within. This relation is most often used to mark when an EVENT is contained entirely by a narrative container.

CONTAINS is a very specific relation implying complete containment within a narrative container, and if you annotate that X CONTAINS Y, it's assumed that there's also an OVERLAP relation between the two. You should only use CONTAINS when you're sure that the nature of the overlap is one of complete containment.

- (213) {March 2005} - Patient underwent [appendectomy]
- {March 2005} CONTAINS [appendectomy]
- (214) [Levaquin] 750 mg p.o. q. day (will [restart] {today})
- {today} CONTAINS [restart]
- (215) [Comparison] is made with prior MRI head [examination] without and with gadolinium from {10-23-03}.
- {10-23-03} CONTAINS [examination]
- (216) An ENT performed the [myringotomy] during {Friday}'s [surgery].
- [surgery] CONTAINS [myringotomy]
 - {Friday} CONTAINS [surgery]
- (217) [Gengraf] 300-mg p.o. b.i.d. ([decreased] in {early June})
- {early June} CONTAINS [decreased]

In addition, we have made one specific regulation involving the use of CONTAINS: All test results or observations are to be linked to the test which generated them (their narrative container) using a CONTAINS relation, *assuming there is actually temporal containment*. This is not always intuitive, because as humans with some inferencing ability, we realize that the tumor likely existed before the CT scan which revealed it, but from a machine-learning perspective, it's important to have that consistency. So, in a section like:

- (218) [Colonoscopy] ({January 7, 2010}): Fair/adequate [prep.], Limited [Colonoscopy] to the distal sigmoid due to an obstructive [lesion]. Diminutive [polyps] of the rectosigmoid,
- {January 7, 2010} CONTAINS [Colonoscopy]
 - [Colonoscopy] CONTAINS [prep.]
 - [Colonoscopy] CONTAINS limited [colonoscopy]
 - [Colonoscopy] CONTAINS [lesion]
 - [Colonoscopy] CONTAINS [polyps]
 - [Colonoscopy] CONTAINS [removed]



Caution!

*There are many relations which seem like a sort of semantic containment (things like part/whole, cause/effect, disorder/symptoms). However, the CONTAINS relation should **only** be used when there exists strict temporal containment (the temporal span of the container fully encompasses those of the contained). This rule will not be violated, and pay attention to DocTimeRel (because a BEFORE EVENT will never CONTAIN an OVERLAP EVENT, etc).*

6.1.3 OVERLAP

OVERLAP is a single temporal relation that encompasses all the different notions of two things happening at the same time, but is less specific than CONTAINS . This can refer to two simultaneous events, an EVENT that occurs during another, larger EVENT or time reference (but where containment is not entirely sure), or any other sense in which two events are occurring in the same timeframe:

- (219) The patient's first [MI] occurred while she was undergoing [chemotherapy].
- [chemotherapy] OVERLAP [MI]
- (220) The patient had some rectal [itching] and mild [pain] {today}, mostly {this morning}.
- {today} CONTAINS [itching]
 - {today} CONTAINS [pain]
 - {this morning} OVERLAP [itching]
 - {this morning} OVERLAP [pain]
- (221) He does have a history of peri-rectal [abscess] with his last round of [chemotherapy].
- [chemotherapy] OVERLAP [abscess]
- (222) She is not [interested] in pursuing chemotherapy> at {this time}
- [interested]_{NEG} OVERLAP {this time}

In short, OVERLAP is meant for situations where two events overlap in some way, but where you're not sure (or don't have enough information to tell) whether there is containment.

OVERLAP is also used for linking TIMEX3s of type SET with other EVENTS, as in (223):

- (223) We'll keep her on rate-control [medications] 100 mg {twice daily}
- {twice daily} OVERLAP [medications]



Caution!

OVERLAP provides relatively little information for the actual processing of text, especially compared to CONTAINS. If you're reasonably sure that the relation is one of containment, you should use CONTAINS instead, and often, one can represent a potential OVERLAP more specifically using narrative containers and a bit of additional thought.

6.1.4 BEGINS-ON

BEGINS-ON signals that the EVENT begins on the EVENT or TIMEX3 it's related to. This type of TLINK will only occur with EVENTS which have a non-trivial temporal span. Relations with punctual EVENTS will usually be marked with BEFORE instead.

- (224) She has had Abdominal [Cramping] since {January}.
- a. [Cramping] BEGINS-ON {January}
- (225) He reports intermittent chest [pain] since his prior [MI]
- a. intermittent chest [pain] BEGINS-ON his prior [MI]

6.1.5 ENDS-ON

ENDS-ON signals that the EVENT ends on the EVENT or TIMEX3 it's related to. As with BEGINS-ON, this type of TLINK will only occur with EVENTS which have a non-trivial temporal span. Relations with punctual EVENTS will usually be marked with BEFORE instead.

- (226) She has had no [bleeding] since her [stitches] were [removed]
- a. [bleeding]_{NEG} ENDS-ON [removed]
- (227) Once intercranial pressure was [decreased], her [seizures] [ceased]
- a. [seizures] ENDS-ON [ceased]
 - b. [ceased] BEGINS-ON [decreased]
 - c. [ceased] TERMINATES [seizures]

Note that ENDS-ON can be used in concert with BEGINS-ON to mark a duration.

- (228) She was on chemo from March through July.
- a. [chemo] BEGINS-ON {March}
 - b. [chemo] ENDS-ON {July}

(Note the ALINK in (228), which links the cessation to the seizure)

6.2 When to TLINK

TLINKs themselves are relatively straightforward, and in fact, the more difficult part of annotating TLINKs is to know when to stop. Without any constraint, one could see making TLINKs between every EVENT in the document, which leads to exponential growth of TLINKs and a tangle of relations which nobody, let alone a machine, would like to unpack.

So, to constrain this process a bit, we have developed five rules to govern TLINKing:

6.2.1 TLINK only when it captures more information than just marking DocTimeRel

Because the 'DocTimeRel' attribute of EVENTS expresses the relation of the event to the time the document or section was written, you will never need to TLINK to the DOCTIME annotations. Marking an EVENT as 'after' in the DocTimeRel field gives us the same information as making an BEFORE TLINK between the EVENT and DOCTIME, so you need not explicitly mark that. Similarly, if one EVENT has a DocTimeRel of OVERLAP and another has a DocTimeRel of AFTER, there's no need to make a TLINK between those two EVENTS.

6.2.2 TLINK all EVENTS to their narrative container, if possible

As previously discussed, most EVENTS will fall into a narrative container of some kind. If a given EVENT is in a narrative container (like "August 22nd" or "during her recovery"), you should always TLINK that EVENT to the TIMEX3 or EVENT which represents that narrative container, using the appropriate link. Once again, though, this should only be done if the result will be more informative than just analyzing the DocTimeRels. See (229), showing an example sentence and the TLINKs required to annotate it:

- (229) {December 19th}: The patient underwent an [EKG] as well as emergency [surgery]. During the [surgery], the patient experienced another [MI], and repeated [tachycardia]
- a. {December 19th} CONTAINS [EKG]
 - b. {December 19th} CONTAINS [surgery]
 - c. [surgery] CONTAINS [MI]
 - d. [surgery] CONTAINS [tachycardia]

In most cases where you have a test and then a series of results, the test itself will form the narrative container for the results, and the day on which the test occurred will itself CONTAIN the test:

- (230) [Colonoscopy] ({January 7, 2010}): Diminutive [polyps] of the rectosigmoid, [removed].
- a. {January 7, 2010} CONTAINS [Colonoscopy]
 - b. [Colonoscopy] CONTAINS [polyps]
 - c. [Colonoscopy] CONTAINS [removed]

Once again, **all test results are TLINKed to the test using CONTAINS, but only if there is actually temporal containment.** Also keep in mind that if the procedure or test has a DocTimeRel of BEFORE, the results will always have a DocTimeRel of BEFORE or BEFORE/OVERLAP.

When annotating, not every EVENT will be a part of a detailed narrative container (more specific than before, after, or during DOCTIME). However, it's vital that you, as an annotator, stop to ask yourself whether each EVENT you examine is a part of a narrative container, and whether the TLINKs you created are sufficient to mark that membership.

6.2.3 TLINK all explicitly stated temporal relations

If the doctor goes out of his or her way to make a temporal statement, a TLINK should be made to reflect that statement. So, if the sentence reads:

- (231) The patient [developed] a [rash] after [treatment]
- a. [treatment] BEFORE [rash]
 - b. [treatment] BEFORE [developed]

You should explicitly mark the treatment as being BEFORE both "developed" and "rash", using a TLINK. Note, though, that if no explicit temporal language is used, no TLINK should be created, and annotator knowledge should not be used to fill these TLINKs in.

6.2.4 Try to only link EVENTS and TIMEX3s within the same sentence

In a perfect world, nearly all TLINKs would occur across two EVENTS or TIMEX3s in the same sentence. That said, often you need to link to an EVENT or TIMEX3 in a previous sentence to put an EVENT in the proper narrative container. In these situations, you may do so, but you should double-check to ensure that there's no other way of going about it, and remember that Coreference annotation will be done to link pronouns and subsequent mentions, so linking an EVENT to a subsequent reference to the narrative container is acceptable as well.

That said, because of the nature of the notes, **TLINKs should never link items in different sections.**

6.2.5 ACTUAL or HEDGED EVENTS should never be linked to HYPOTHETICAL or GENERIC EVENTS, and vice versa.

This may seem quite specific, but HYPOTHETICAL and GENERIC EVENTS aren't really on the same timeline as other EVENTS dealing with the patient's actual care. Because of this, it can be quite tricky to link them to the overall timeline, and we need to take care to avoid having non-real EVENTS showing up on patient timelines. To avoid the complications and potential broken relations when non-real EVENTS are pruned from timelines, ACTUAL or HEDGED EVENTS should never be linked to HYPOTHETICAL or GENERIC EVENTS, and vice versa. In this way, "real" EVENTS are never linked to non-real ones.

(232) Adjuvant [chemotherapy] following her upcoming [surgery] would generally be recommended, but given her poor [health], this is not an option.

Here, because the [chemotherapy] is HYPOTHETICAL, it cannot be TLINKed to [surgery], even given the explicit mention. Further discussion can be found in Section 11.3.

6.2.6 You do not need to TLINK TIMEX3s to one another

Although there is certainly a temporal relation between, say, "January 15th, 2009" and "March 2013", part of the post-processing of these annotations is the normalization of these Temporal Expressions, which will allow us to order these expressions on a timeline based on the times they represent. Although you will certainly still TLINK EVENTS and TIMEX3s, the TIMEX3s in the document can almost always be temporally ordered without the annotator's help.

6.3 Expressing TLINK Types in Point Algebra

A more precise (although perhaps more esoteric) way to express temporal relations is using point algebra. In the style of Allen 1983 ([?]), we can take the start and end of event A to be A- and A+, respectively, and compare those points to the start and end of EVENT B, B- and B+, using < ("less than") and > ("greater than") to indicate position on the timeline.

Given this notation, our TLINK types can be expressed as below:

Relation	Notation	Meaning
BEFORE	A+ < B-	"A is before B"
CONTAINS	(A- < B-) AND (A+ > B+)	"A contains B"
OVERLAP	(A- < B- < A+) OR (B- < A- < B+)	"A and B overlap"
BEGINS-ON	A+ = B-	"B begins-on A"
ENDS-ON	A- = B+	"B ends-on A"

Note that our definition of OVERLAP is symmetrical, in that "A OVERLAP B" and "B OVERLAP A" mean the same thing, and which EVENT starts first is not recoverable from the annotation without prior knowledge of both A- and B-.

6.4 Tricky TLINK examples

This section includes a variety of examples showing tricky situations which come up when TLINKing. The first, (233), shows us both narrative container nesting and reference to a previous study:

- (233) A subsequent [MRI] on {November 17} [demonstrated] significant [reduction] in the anorectal wall thickening measuring 0.9-cm from a previous [1.7-cm] in {July 2009}.
- {November 17} CONTAINS [MRI]
 - [MRI] CONTAINS [demonstrated]
 - [MRI] CONTAINS [reduction]
 - {July 2009} CONTAINS [1.7-cm]
 - [1.7-cm] BEFORE [reduction]

Here's a case where inference is both easy and unwarranted:

- (234) The [cancer] was [resected] by Dr. Collins on the 15th.
- [cancer] OVERLAP [resected]
 - {the 15th} CONTAINS [resected]
 - [cancer] OVERLAP {the 15th}

Many annotators want to mark that [cancer] ENDS-ON [resected]. Alas, if "resected" meant cancer ENDS-ON resection, it would be a wonderful world, but that's not an inference we can safely draw. In this case, unless it's qualified ("We resected the cancer, and since margins were clear, we can say the cancer is gone"), we must limit ourselves to saying that the resection occurred in its container, and that it overlaps the cancer.

7 ALINK Annotation

The final sort of annotation performed on the data is the ALINK (or "aspectual link") annotation. ALINKs are created between an aspectual EVENT and a non-aspectual EVENT. Any EVENT previously marked with the class ASPECTUAL will be ALINKed to another, non-aspectual event, and you will never make an ALINK which includes a TIMEX3 or two non-aspectual EVENTS. ALINKs only occur when you have an aspectual EVENT, and are much less common in the text than TLINK annotations. Relatively few EVENTS will have an associated ALINK.

Unlike TLINKs, no matter the circumstance, an ALINK should never cross a sentence boundary.

As with TLINKs, they are created interactively, but have the basic form:

- (235) EVENT1 [aspectual relation] EVENT2

7.1 ALINK sub-types

There are four different aspectual relations used in the schema, CONTINUES, INITIATES, REINITIATES, and TERMINATES.

7.1.1 CONTINUES

CONTINUES is used when an aspectual event shows the continuation of another event:

- (236) We will [continue] to [monitor] LFTs carefully along with his weight.
 a. [continue] CONTINUES [monitor]
- (237) The patient will [remain] on [dialysis] until her condition [changes].
 a. [remain] CONTINUES [dialysis]
 b. [remain] ENDS-ON [changes]
- (238) She is not interested in pursuing chemotherapy at this time but is interested in [continued] [surveillance]
 a. [continued] CONTINUES [surveillance]
- (239) We will [continue] to [monitor] heart rate and rhythm along with serial cardiac markers and electrocardiograms to rule her out for any cardiac involvement
 a. [continue] CONTINUES [monitor]

7.1.2 INITIATES

INITIATES is used when an aspectual event indicates the start or initiation of another event:

- (240) Patient will [begin] a high-fiber [diet] upon [release].
 a. [begin] INITIATES [diet]
 b. [begin] BEGINS-ON [release]
- (241) We will [start] Ms. Miller on a normal saline [infusion] at 75 an hour for a total of 1 L
 a. [start] INITIATES [infusion]

7.1.3 REINITIATES

REINITIATES is used when an aspectual event indicates that another event will be restarted or reinitiated:

- (242) [Levaquin] 750 mg p.o. q. day (will [restart] {today})
 a. [restart] REINITIATES [Levaquin]
 b. {today} CONTAINS [restart]
- (243) His anterior chest [rash] has not [reoccurred] since the [PCN] VK was [discontinued] {24-hours ago}.
 a. [reoccurred]_{NEG} REINITIATES [rash]
 b. [discontinued] TERMINATES [PCN]
 c. [discontinued] BEFORE [reoccurred]_{NEG}
 d. [discontinued] OVERLAP {24-hours ago}

7.1.4 TERMINATES

TERMINATES is used when an aspectual event indicates the ending of another event:

- (244) Because of this [reaction], [Allegra] will be [discontinued]
 a. [discontinued] TERMINATES [Allegra]
- (245) We will [hold] her [heparin] until after the [surgery].
 a. [hold] TERMINATES [heparin]
 b. [hold] ENDS-ON [surgery]
- (246) Patient [nausea] was successfully [stopped] by 1-mg [Ativan] p.r.n.
 a. [stopped] TERMINATES [nausea]



Caution!

ALINKs are much less common than TLINKs, usually only 1 or 2 per document, but they're no less important and easy to overlook. Remember that any time you have an EVENT which has the type ASPECTUAL, you'll need to create at least one ALINK.

8 DOCTIME and SECTIONTIME Annotation

Because all of these annotations are linking the EVENTS and TIMEX3s to the time of patient service, it's important that we specify what that time actually is. Each note should include a line at the very start or end of the document which states the service time, and we select that time as the DOCTIME. DOCTIME can be considered a special sort of TIMEX3, and is annotated similarly.

It's worth noting though, that because of the complexities of medical record extraction, some sections (separated by [start section id="xxxx"] blocks) might have a separate, local time. These are often vital signs and medication sections, and they will always include a specific date, which will be marked not with DOCTIME, but with SECTIONTIME. In these cases, any DocTimeRel annotations will link to this, section-specific time. *If no separate date is explicitly given for a section, assume that it shares the overall DOCTIME.*

So, at the top of the document, you'll find a line like the below:

(247) [head start date="[12/13/2010]" rev date="12/13/2010" rev="0002"]

Here, the [12/13/2010] would be marked as DOCTIME, as "start date" is the closest thing to the "Document Time" that we have in the Colon Cancer Corpus, but other corpora may well have different dates and titles for them.

Sometimes, as discussed before, a specific section has an explicit, different timestamp than the rest of the document:

(248) [start section id="20104"] Alendronate FOSAMAX 70 mg tablet 1 tablet by mouth every week. Colace 100-mg capsule 1 capsule by mouth two times a day. These are the patient's medications as of [Saturday, February 20, 2010 at 3:34 PM]. [end section id="20104"]

Here, [Saturday, February 20, 2010 at 3:34 PM] would be considered SECTIONTIME, and would be annotated as such.

9 Annotating Pathology Notes

Pathology notes, in our schema, are mostly interesting in that little temporal information is actually contained within them. Although there are no major deviations from the schema as described previously, this section will discuss how to properly annotate a pathology note when you are asked to do so.

Here is a sample pathology note:

```
[meta rev_date="2/24/2010" start_date="2/25/2010" rev="0001"]
```

```
[start section id="$final"]
```

Final Diagnosis

A. Colon, right hemicolectomy: Invasive grade 4 (of 4) adenocarcinoma, forming a 6.8 x 6.3 x 1.6 cm ulcerative, necrotic, annular mass in the cecum involving the ileocecal valve. Tumor infiltrates through the muscularis propria to involve the pericolic fat and the free serosal surface (visceral peritoneum). Lymphatic/vascular invasion is present. The appendix shows fibrous obliteration of lumen. The surgical resection margins are negative for tumor. Multiple (2 of 33) regional lymph nodes are positive for metastatic adenocarcinoma. [AJCC pT4N1]

This final pathology report is based on the gross/macroscopic examination and the frozen section histologic evaluation of the specimen(s). Hematoxylin and Eosin (H&E) permanent sections are reviewed to confirm these findings. Any substantive changes identified on permanent section review will be reflected in a revised report.

```
[end section id="$final"]
```

```
[start section id="$gross"]
```

Gross Description

A. Received fresh labeled ""right colon"" is a right hemicolectomy specimen consisting of 10.7 cm of terminal ileum, 23.2 cm of cecum and ascending colon and a 9.2 x 0.5 cm appendix. A 6.8 x 6.3 x 1.6 cm ulcerative, necrotic, annular mass is present within the cecum and involving the ileocecal valve and terminal ileum. Tumor perforation is absent. The mass is located 10.2 cm from the proximal mucosal resection margin and is present at the free serosal surface. There is 1.8 cm of fat over the tumor posteriorly. Multiple (33) lymph nodes are identified within the mesenteric fat. Representative sections are submitted. Grossed by KING.

```
[end section id="$gross"]
```

In this note, a few things are apparent. First, the majority of EVENTS will have a DocTimeRel of OVERLAP, as the pathologist is speaking only to the nature of the sample that he or she has been given, at the time it is available. Although it is true that "Gross Description" and "Final Diagnosis" were written at two different times, remember that the DocTimeRel attribute marks EVENTS timing relative to the writing of each specific section, where that may differ from the overall document time.

Because of all these factors, in this entire note, only six EVENTS have a DocTimeRel which is not OVERLAP:

- Right Hemicolectomy - BEFORE
- gross/macroscopic examination - BEFORE
- frozen section histologic evaluation of the specimen(s) - BEFORE
- revised report - AFTER
- received - BEFORE

- permanent section review - AFTER

These notes are very formulaic, and thus, these non-OVERLAP EVENTS are usually very easily identified.

It's worth noting as well that in this pathology note (as in most of them), there are no TLINKs to be made. Here, none of the EVENTS have explicit temporal relations defined, and all implied relations are captured by DocTimeRel. Thus, all that would be necessary to fully annotate this note is the marking of all EVENTS, marking of the properties for each, and a cursory scan to make sure the note you're working with doesn't have any odd TLINKs.

9.1 AJCC Staging Codes

One EVENT which we **must** capture is the AJCC staging code, which is perhaps the most critical part of the pathology note for reviewing physicians. Although not every note has one, when present, this code will always be in the "Final Diagnosis" section", and is usually relatively straightforward to find, as it will occur on its own, or occasionally in a larger introductory sentence. It will not always include the characters "AJCC" in the span, and you are only wanting to capture the code itself, rather than the descriptors. Below are several examples of this code in context:

- (249) Multiple (2 of 33) regional lymph nodes are positive for metastatic adenocarcinoma. AJCC [pT4N1]
- (250) Multiple (3 of 12) regional lymph nodes are involved by metastatic adenocarcinoma. AJCC [pT3N1MX]
- (251) With available surgical materials, AJCC [pT3N0] (7th edition, 2010).
- (252) A separate tubular adenoma with low grade dysplasia is also identified. AJCC, 7th edition [pT3N2bMX]
- (253) AJCC [pT3N0MX].

No matter its form, please try to capture the AJCC staging code in any notes you annotate as EVENTS with a DocTimeRel of OVERLAP

9.2 Guidelines for Annotating Pathology Notes

Although you should always keep your eyes open for unusual cases, most notes can be annotated quickly according to the below guidelines:

1. Mark all EVENTS using the general guidelines above (or, in many cases, check all pre-annotated EVENTS for accuracy).
2. Mark all EVENTS as OVERLAP, but...
3. Look out for the below EVENTS which generally have a DocTimeRel other than OVERLAP:
 - Mention of the procedure which produced the sample ("Right hemicolectomy"), always BEFORE
 - Mention of the gross description procedures in the final diagnosis section ("frozen section histologic evaluation of the specimen(s)", "gross/macrosopic examination"), always BEFORE
 - The often-promised "Revised Report", always AFTER
 - Mention of receiving the specimen in the Gross Description, always BEFORE
4. Ensure that the final diagnosis is captured as an EVENT ("Invasive grade 4 (of 4) adenocarcinoma", for instance)
5. Ensure that the AJCC Diagnosis and Staging code is captured as an EVENT

6. Check to make sure you don't have any TLINKs in the document
 7. Mark the DOCTIME at the top of the note ("meta rev_date")
- Once these steps are completed, your pathology note should be fully annotated.

NOT FOR REDISTRIBUTION

10 Schema Description

In an effort to clarify the schema at a glance, here's a quick summary of the Clinical Temporal Relations Annotation Schema, using the structure described below. Default values, where applicable, are in ***bold italics***:

- **Entity/Relation Type**
 - Property of that entity/relation type
 - * Choice 1 for that property
 - * Choice 2 for that property
 - * ...
- **EVENT**
 - DocTimeRel
 - * BEFORE
 - * OVERLAP
 - * AFTER
 - * BEFORE/OVERLAP
 - Type
 - * *N/A*
 - * ASPECTUAL
 - * EVIDENTIAL
 - Polarity
 - * ***POS***
 - * NEG
 - Degree
 - * *N/A*
 - * MOST
 - * LITTLE
 - Contextual Modality
 - * ***ACTUAL***
 - * HYPOTHETICAL
 - * HEDGED
 - * GENERIC
 - Contextual Aspect
 - * *N/A*
 - * NOVEL
 - * INTERMITTENT
 - Permanence
 - * FINITE
 - * PERMANENT
 - * ***UNDETERMINED***
- **TIMEX3**
 - Class
 - * DATE
 - * TIME
 - * DURATION
 - * QUANTIFIER
 - * PREPOSTEXP
 - * SET
- **TLINK**
 - Event
 - * (This is a slot, designed to be filled with the EVENT being TLINKed)
 - Related to
 - * (This is a slot, designed to be filled another EVENT or TIMEX3 to which the first EVENT is being temporally linked)
 - Type
 - * BEFORE
 - * OVERLAP
 - * CONTAINS
 - * BEGINS-ON
 - * ENDS-ON
- **ALINK**
 - Event
 - * (This is a slot, designed to be filled with the EVENT (of type "ASPECTUAL") being ALINKed)
 - Related to
 - * (This is a slot, designed to be filled another EVENT to which the first EVENT is aspectually linked)
 - Type
 - * CONTINUES
 - * INITIATES
 - * REINITIATES
 - * TERMINATES
- **DOCTIME**
- **SECTIONTIME**

10.1 Default Values

Some of the attributes of EVENTS, TLINKs, TIMEEX3s and ALINKs have default values, which will automatically be entered by the annotation tool. This is a list of those default values:

Default Values for EVENT

Property	Default Value
EVENT.type	(N/A)
EVENT.DocTimeRel	Must be specified
EVENT.polarity	POS
EVENT.degree	(N/A)
EVENT.contextualaspect	(N/A)
EVENT.contextualmodality	ACTUAL
EVENT.permanence	UNDETERMINED

Default Values for TIMEEX3

All values for TIMEEX3 annotations must be specified for each instance

Default Values for TLINK

All values for TLINK annotations must be specified for each instance

Default Values for ALINK

All values for ALINK annotations must be specified for each instance

11 Special Cases and Constructions

11.1 Avoiding Inference

One thing to keep in mind when annotating is that you, being human with even basic understanding of the world and of the way that medicine work, will easily be able to infer relations that are simply not present in the text. Take, for example, the below sentence:

(254) We [diagnosed] her [cancer] last week.

Here, it's very tempting to want to assign the cancer and the diagnosis different temporal realities, to claim that the cancer preceded the diagnosis (as well as last week), and that the diagnosis was simply one point in a longer history of cancer.

However, we don't really know that the cancer preceded the diagnosis, and it's really human inference that gives us that information. It is possible, however remotely, that the cancer could have existed since birth, or perhaps cropped up 5-10 minutes before the biopsy. Regardless, we know that the diagnosis and cancer OVERLAP, but we don't know much else.

Diagnoses (and notes in general) are a flashbulb in a dark room. We can infer that what we see during the flash was there before and persists after, but that's not a level of surety at which we can annotate, unless the doctor is more explicit ("we diagnosed her with carcinoma in Fall 2007, but it was there for many years prior").

Compare this to the below:

(255) We [observed] her [seizure] last week.

In this case, we again have an EVIDENTIAL EVENT alongside a regular one, but here, it would be silly to infer that the seizure has a different temporal timeframe than then observation. But note that nothing has changed syntactically or linguistically to indicate that difference, so this would be a very difficult point to teach a non-human. As such, although we assume there's a different timeframe for the diagnosis and the cancer itself, we should not annotate based on that assumption.

Similarly, you'll see sentences like:

(256) On the 12th, she underwent an [EKG] and emergency [bypass] following an [MI].

As humans, we understand that the test (the EKG) likely occurred before the invasive procedure (the bypass). But when we compare:

(257) On the 12th, she underwent an [MRI] and [CT] following a car [accident].

We see that there is no easily understood order between [MRI] and [CT], and because there's no linguistic difference, just human understanding, we should not annotate the inferred relation between [EKG] and [bypass].

This may seem like a silly point to belabor, but recall that the eventual goal is to use this data to train software, and that software will have no understanding of this distinction. As such, even when it seems like you're throwing away easily recoverable information, you should always know your audience and annotate based on the text, rather than based on your understanding of the world.

11.2 CONTAINS and NEGATIVE Polarity EVENTS

One commonly asked question is whether negative EVENTS can be CONTAINED at all. For instance:

(258) Vaginal [hysterectomy] without [oophorectomy], 1994

Arguably, she is always not having oophorectomies, except when and if she ever does have one.

Pragmatically speaking, we don't temporally enjoin negated activities unless relevant or exceptional. If one asks a traveler about where they are and they reply "Not in Washington", the implication is that they should be in Washington, but due to whatever unpleasant or unexpected situation, they're not. Thus, the current state of not being in Washington is mentionable, exceptional, and specific, whereas in the rest of their life up until that point, they hadn't felt the need to bring up the fact that they're not in Washington.

Similarly, if a note says "she didn't have oophorectomies in 1994", we argue that it represents an exceptional (and thus, mentionable) state of not currently being de-ovaryed during the mentioned period, and that exceptional state itself is temporally contained within 1994.

Thusly, we argue that negated EVENTS CAN be temporally contained by virtue of the specificity and pragmatic exceptionality of their mention, and in this case, 1994 CONTAINS [oophorectomy].

11.3 Discussions, Generalizations, and other EVENTS which aren't on the timeline

One of the biggest sources of difficulty for annotators is the presence of EVENTS which don't really belong on the patient's timeline per se, but are present in the notes. Let's use the below examples to discuss.

- (259) We're going ahead with adjuvant [chemotherapy] for Mrs. Parks, starting next week.
- (260) I briefly discussed the [risks] of adjuvant [chemotherapy] with Mr. Peloquin, along with those of several other treatment options.
- (261) Adjuvant [chemotherapy] following [surgery] is generally recommended in situations similar to this.

In all cases, clearly, [chemotherapy] represents an EVENT, namely, the administration of chemotherapy as a secondary measure following a primary treatment. However, only in (259) is the chemotherapy actually on the patient's timeline, something which is presumed to occur. In (260), the chemotherapy will not necessarily occur for the patient, and in (261), we have a generalization about patient care, rather than a specific discussion of that patient's treatment plan. These EVENTS are clearly not a part of the patient's actual course of treatment, and thus shouldn't show up on his or her timeline. If the note reads "We discussed her upcoming surgery as well as the possibility for future recurrence", we certainly don't want "recurrence" showing up on a factual timeline, but we certainly would want "surgery" to show up.

In order to properly train our systems to find EVENTS, all EVENTS must be marked, however, we must make sure that we can reconstruct the reality (and the applicability to the timeline) of all EVENTS. Our primary means for handling this is the Contextual Modality property of EVENTS. As discussed in Section 3.7, we use HYPOTHETICAL and GENERIC to mark EVENTS which, although certainly events, don't belong on the timeline. As you may have already guessed, in (260), "adjuvant chemotherapy" is a HYPOTHETICAL EVENT, and in (261), it's a GENERIC EVENT.

Using these two modalities, you should be able to annotate all of the cases where EVENTS are discussed which do not belong on the timeline (or belong on a hypothetical fork of the timeline).

11.3.1 Ground Rules for annotating verbs of discussion

As discussed previously, these EVENTS, by virtue of their modality, aren't really on the same timeline as other EVENTS dealing with the patient's actual care. Because of this, it can be quite tricky to link them to the overall timeline. As such, we have a few rules, all previously discussed, which you will find particularly useful:

1. If an EVENT is HYPOTHETICAL, mark DocTimeRel as best you can using the same guidelines you would for all other EVENTS, as if they were actual EVENTS.
2. If an EVENT is GENERIC, and thus, a statement of fact, DocTimeRel will always be OVERLAP
3. Because these EVENTS are on timelines of their own, you should never TLINK an EVENT marked HYPOTHETICAL or GENERIC to an ACTUAL or HEDGED EVENT.
4. In situations with a verb of discussion or talk, no matter the modality of what was discussed, the verb itself will have a modality of ACTUAL, as the discussion really did take place.

11.3.2 Examples of discussion

- (262) We [discussed] that Adjuvant [chemotherapy] following [surgery] is generally recommended in situations similar to this.
- a. [surgery] BEFORE [chemotherapy]

Here, we have a straightforward discussion of GENERIC EVENTS. So, the EVENTS are as marked above, with all three having an OVERLAP DocTimeRel. In addition, we'd TLINK [surgery] BEFORE [chemotherapy], both GENERIC EVENTS. [discussed] is ACTUAL, as the discussion actually occurred.

- (263) I briefly [mentioned] the risks of adjuvant [chemotherapy] with Mr. Peloquin, along with those of several other treatment options.

Here, the risks of a hypothetical course of chemotherapy are mentioned. [mentioned] would be ACTUAL, with a DocTimeRel of OVERLAP, and adjuvant [chemotherapy] would be hypothetical.

(264) We also [discussed] the risk of wound [infections], urinary tract [infections], and [pneumonia].

Again, ACTUAL discussion, with three hypothetical EVENTS.

(265) The risk of the [anastomosis] was the primary risk [discussed] including the possibility of an anastomotic [leak], need for [intervention], possibly even [surgery] and a temporary [stoma]

Don't let the rearrangement of the sentence fool you, this is just like the prior examples. [discussed] would be ACTUAL, and all the other EVENTS would be hypothetical. Although there are temporal relations among the EVENTS here, they're not explicitly stated in the document, so they shouldn't be marked.

(266) I am [concerned] about use of adjuvant [chemotherapy] such as [FOLFOX] in this elderly man
a. [FOLFOX] OVERLAP adjuvant [chemotherapy]

Although there's no explicit discussion here, the doctor is still making a statement. Here, you'd have an ACTUAL [concerned], HYPOTHETICAL [chemotherapy] and [FOLFOX], and an OVERLAP TLINK between [FOLFOX] and [chemotherapy], both hypothetical.

11.4 Family History/History of..

Patient history and family history are treated as narrative container EVENTS in our schema. So, in the sentences:

(267) Past medical [history]: Common variable [immunodeficiency]
a. [history] CONTAINS [immunodeficiency]

We have two EVENTS. [medical history] is an EVENT whose DocTimeRel is BEFORE-OVERLAP (because the patient's medical history started before the document time, but continues through the note). [immunodeficiency] is also an EVENT, with a DocTimeRel of BEFORE (as the patient is not experiencing this immunodeficiency at the time of writing). The TLINK between Common variable immunodeficiency and medical history simply indicates that the disorder occurs as part of the medical history narrative container, and because we're treating history as a persistent state, during the patient's medical history. As such, if the sentence read "Common variable immunodeficiency and diabetes", Diabetes would also be an EVENT, and would also be TLINKed into the "medical history" narrative container. Here's another example of this approach:

(268) No [history] of [cough] or aspiration [pneumonias]
a. [history] CONTAINS [cough]_{neg}
b. [history] CONTAINS [pneumonias]_{neg}

Family history works similarly:

(269) Family [history] Mother and father with allergic [rhinitis], siblings with allergic [rhinitis] and atopic [eczema]
a. [history] CONTAINS [rhinitis]
b. [history] CONTAINS [eczema]

Once again, "family history" is an EVENT, DocTimeRel BEFORE-OVERLAP, and each of the disorders described are EVENTS with DocTimeRel BEFORE. The TLINKs mark containment of the various disorders within the abstract "family history".

When there is no explicit usage of "history" or "family history", mark the historical EVENTS, giving the proper DocTimeRel, even though they're not actually issues with the patient. We're training our system to find EVENTS, and those things are definitely EVENTS, even though they've got a different meaning to us humans.

11.4.1 Medical History sections

Often, you'll find sections labeled "medical history", listing a variety of disorders:

(270) History of diverticulitis, Asthma, History of cutaneous sarcoidosis, Type II diabetes mellitus, adult onset, using insulin

In general, the way a given EVENT is expressed provides temporal information. If a history item is expressed simply as "Asthma", this implies that it is an ongoing condition (DocTimeRel BEFORE/OVERLAP), compared the others, described as "history of", which are likely treated or no longer a factor in patient treatment (DocTimeRel BEFORE). Usually, no TLINKs are possible, but DocTimeRel most certainly is.

However, here, you may need to cross-reference the other portions of the note to determine the exact DocTimeRel and Narrative Container structure of the EVENTS in the history section.

11.5 Pre- and Post- expressions (preoperative, post-treatment, etc)

The adjectives "preoperative" and "postoperative" (as well as related terms like "post-surgical", "post-injury", "pre-treatment", "post-partum", "intraoperatively", "post-prandial") and present a particular difficulty for our schema when used without explicit mention of their referents, as they simultaneously express two inseparable temporal meanings: first, they introduce an EVENT (like an operation or an injury), and second, they tell us that the noun they modify occurred either before or after this operation.

Because "preoperative" and "post-treatment" (and their ilk) actually denote temporal spans ("everything until the surgery", "any time after the treatment"), they're actually a sort of TIMEX3. So, all of these pre- and post-expressions will be marked as being TIMEX3 of the type "prepostexp" ("Pre- Post- Expression"), which function much like other narrative container anchors.

First, an easy example:

(271) Patient underwent a partial [hemicolecotomy] in {July 2009}. {Postoperative} [scarring] [noted] during [exam].

- a. {July 2009} CONTAINS [hemicolecotomy]
- b. [hemicolecotomy] BEFORE [scarring]
- c. [hemicolecotomy] BEFORE [noted]
- d. [hemicolecotomy] BEFORE [exam]
- e. [hemicolecotomy] BEFORE {Postoperative}
- f. [exam] CONTAINS [noted]
- g. {Postoperative} CONTAINS [scarring]

Here, you can see that there's a large amount of temporal information expressed here, but most important is the [postoperative] TIMEX3 and the two associated TLINKs. The scarring is TLINKed to postoperative with CONTAINS because the scarring occurs during the span of time which "postoperative" designates. Then, we mark that postoperative BEGINS-ON the surgery, because the surgery's end represents the start of the postoperative period.

This is relatively straightforward, but there's not always an explicit reference to the surgery in the note:

(272) The patient exhibits postoperative changes.

Here, we have no EVENT for the operation which we can TLINK the "changes" to. In this situation, "postoperative" would still be a TIMEEX3, and the changes would OVERLAP it.

(273) The patient exhibits {postoperative} [changes].
 a. {postoperative}_{prepostexp} CONTAINS [changes]

A second example of this would be:

(274) Patient is in [recovery], no {post-operative} [nausea].
 a. {post-operative}_{prepostexp} CONTAINS [nausea]

Once again, post-operative here stands in for the span of time following the surgery, allowing us to capture that the lack of nausea occurred following the unmentioned procedure.

(275) The patient's {preoperative} health is [good].
 a. {preoperative}_{prepostexp} CONTAINS [good]

Example (275) is a bit more opaque, but preoperative in this case stands in for the span of time before the surgery it introduces which presumably will occur in the future (given the tenses in the sentence). So, "preoperative" will be an EVENT, which will OVERLAP the document time. The "health is good" EVENT is occurring at the document time, so it has a DocTimeRel as OVERLAP, and then finally the TLINK shows us that the preoperative period CONTAINS the good health. Here's a slight variation:

(276) The patient's {preoperative} [health is good], and {next week}'s [tonsillectomy] should proceed without difficulty
 a. {preoperative} CONTAINS [good]
 b. [good] BEFORE [tonsillectomy]
 c. {next week} CONTAINS [tonsillectomy]
 d. {preoperative} ENDS-ON [tonsillectomy]

With the explicit addition of the referent in (276), we would TLINK "health is good" to "tonsillectomy" (BEFORE) in addition to the OVERLAP link to the PrePostExp. The majority of cases involving pre- and post- expressions will include referents and thus should be relatively transparent to annotate, but in the event of a missing referent, the above guidelines should allow you to capture the temporal information implied.

Finally, it's critical to note that words marked PREPOSTEXP will not always begin with "pre-" or "post-". Terms like "intraoperatively" can sometimes be PREPOSTEXP as well, as in (277) below:

(277) [Intraoperatively], there were no difficulties securing his airway. He received 4 liters of crystalloid, made 300 mL of urine and estimated blood loss was 50 mL. He received a dose of DDAVP in the OR out of concern for uremic platelet dysfunction

11.6 Annotating the word "plan" or "plans"

Plan is a temporally difficult word. In some cases, it's clearly temporally relevant:

- (278) I'd discussed that she'll need to come in for the EGD with biopsies. followed by a CT scan. We're hoping to implement this plan in the next few weeks.

In this case, it's clearly an EVENT as it needs to be implemented at a specific time and TLINKed with 'next few weeks'. But in some cases, its status as an EVENT is more ambiguous.

- (279) I have explained this plan to both Miss Mullins and the primary team.

Here, there's no strong TLINK, and the plan itself isn't necessarily temporally bound.

To avoid this ambiguity and improve annotator agreement, we've decided that **in all cases, "plan" will be considered an EVENT**. Do your best to annotate DocTimeRel and all other EVENT properties, making TLINKs which may be necessary, but no matter the usage, we should always be marking 'plan' (or 'plans') as an EVENT. This rule applies to both verbs and nouns, where relevant.

11.7 Annotating "prior" and "at the same time"

"Prior", "at the same time", and other related temporal indicator words or phrases are often complicated in their role within the schema, and as such, merit special discussions.

11.7.1 Annotating "Prior"

The most common mistake made by new annotators is to consider "prior" itself to be a TIMEX3. It is, of course, a word carrying considerable temporal information, but it's not a TIMEX3. Prior is a marker of a temporal relationship, like the word "before" or "during", rather than an explicit reference to a time or date (like "Yesterday", "Last weekend" or even "at 2pm"). As such, the word "prior" itself will never be a TIMEX3, although, as shown below, it will occasionally show up as part of a longer temporal expression.

The best way to illustrate the varied uses of "prior" in the data is to discuss a series of examples:

- (280) Healthy prior to admission and only co-morbid conditions.
a. [healthy] BEFORE [admission]

Here (as in many cases), "prior" is marking temporal relationship, here stating that the [healthy] comes BEFORE [admission]. This is the case with most usages, where "prior" simply indicates a temporal link ("She should meet with me prior to discharge", "The electrolyte panel should be reviewed prior to the consult"). This link is usually of type BEFORE, but occasionally marks ENDS-ON ("She had no symptoms prior to her coughing"). Note that in this example (as with most), the word "prior" itself is left unmarked.

- (281) Prior to her [MI], she experienced no [signs] of cardiovascular illness"
a. [signs]_{NEG} ENDS-ON [MI]

Example (281) is identical to (280), except for a syntactic change pulling the "prior to her MI" to the front of the sentence. Once again, prior indicates the given TLINK.

- (282) [Omeprazole] should be taken 1 hour prior to [meals].
a. {1 hour} OVERLAP [Omeprazole]

Note that here, "1 hour prior to meals" is itself a temporal expression (which happens to include the term "prior"), to which Omeprazole is linked. No further links are possible in this schema.

(283) Prior [treatment] with Prilosec was not felt to be [effective].

Here, all that "prior" is giving us is that the treatment with Prilosec has a DocTimeRel of "before". No explicit TLINK is given or indicated. This usually the case when "prior" is being used as a modifier ("Her prior illness", "the prior surgery", etc).

11.7.2 Annotating "at the same time"

Similarly to "prior", "at the same time" (and derivative expressions) is not a TIMEX3, but a marker of a temporal relationship. For example:

(284) We'll order the EGD as a complex procedure so that Botox injection can be done at the discretion of the endoscopist at that same time.
a. [EGD] CONTAINS [injection]

Here, "at that same time" is indicating that [EGD] can CONTAIN [injection], and is itself unmarked. This is the most common usage, marking OVERLAP or CONTAINS TLINK relations.

(285) She had a hysterectomy with scar ovaries removed and appendix removed at the same time
a. [hysterectomy] OVERLAP [removed]

In (285), again, "at the same time" is left unmarked.

(286) I met with the patient at 2pm today, and told her I'd check in at the same time tomorrow
a. at {the same time tomorrow} CONTAINS [check]

(286) shows one of the relatively rare uses of "at the same time" as a TIMEX3. Here, {the same time tomorrow} is a TIMEX3, which CONTAINS [check].

(287) There is good evidence that this procedure will improve her condition, but at the same time, it could easily lead to additional complications.

This usage is a red herring, as this "at the same time" is simply a turn of phrase used to separate opposing statements, rather than indicating a worthwhile temporal link.

12 Non-annotated sections of clinical notes

Not all portions of the clinical note should be temporally annotated. The following sections, given both with a Mayo section descriptor and a qualitative description (where applicable), should not be annotated at all (for a variety of reasons). These sections should not be annotated even if not labeled with the section id given below, or if they're included in another section (as "Patient Education" and "Advance Directives" often are). That said, if you're not sure whether a given section of a note should be included, contact your annotation supervisor.

- 20104 - Current Medications
- 20105 - Allergies
- 20116 - Advance Directives
- 20138 - Patient Education
- "Patient Diet"

12.1 Duplicated Text

Some notes have significant amount of copy/pasted text in them, generally legal disclaimers or patient information. Below is a list of known oft-duplicated text sections. If, as you are annotating, you come across a section or paragraph which is identical to one listed below, you should **not** annotate with EVENT, TIMEX3, or TLINKs, and instead should select the entire paragraph or section which is copy-pasted and mark with DUPLICATE.

- The patient has a surgical incision, and should monitor for signs and symptoms of infection which include, but are not limited to, redness at the site of the incision, purulent drainage from the incision, separation of the wound edges, nausea, vomiting, general malaise, fevers greater than 101.5 degrees Fahrenheit, and chills. If these occur in combination or if there is gross purulence coming from the wound, the patient should call.
- The surgical wound should be kept clean, and this is best accomplished by letting shower water run across it. The wound should not be directly rubbed or scrubbed for a period of at least one month. The wound should not be immersed in water (i.e. hot tub, bath tub, swimming, etc.) for a period of six weeks.
- Your incision has been closed with sutures. Sutures should be removed 10-14 days after surgery. You may schedule an appointment with your primary care provider to have the sutures removed. Your provider or their nurse may remove the sutures.
- Take your pain medications as instructed. It is best to take pain medications before your pain becomes severe. This will allow you to take less medication yet have better pain relief. For the first 2 or 3 days it may be helpful to take your pain medications on a regular schedule (e.g. every 4 to 6 hours). This will help you to keep your pain under better control. You should then begin to take fewer medications each day until you no longer need them.
- Do not consume more than 4000 mg or 4 gm of Tylenol or acetaminophen-containing products (Percocet, Vicodin, etc.) within a 24 hour time period to avoid damage to your liver.
- Do not take pain medication on an empty stomach. This may lead to nausea and vomiting.
- Do not drive, operate heavy machinery, ride motorcycles or ATVs, drink alcohol, or take other drugs that make you tired or sleepy while taking narcotic pain medications (Percocet, Vicodin, oxycodone, etc.). Do not participate in any dangerous activity while taking pain medications. Pain medications may decrease your ability to make safe decisions.
- If you currently smoke cigarettes, cigars, or pipes; chew tobacco; or have done so in the past 12 months, it is important to quit. Talk with your physician about different ways to stop smoking. For additional information about smoking cessation, go to <http://www.flickr.com/fluinfo.php> or call 234-860-4879.
- Narcotic pain medications can cause numerous side effects including nausea, constipation, and confusion. You should not operate a motor vehicle or consume alcohol while taking narcotic medications. The medication may be used for short time after surgery. At any time, Tylenol can be used instead. 1000 mg of Tylenol can be taken safely in most patients every six hours. While taking narcotic, you should also take a stool softener such as Colace, 100 mg twice daily to prevent constipation. You may take Percocet or Vicodin up to 2 tablets every 4 hours. Acetaminophen is the active ingredient in Tylenol, found in most narcotic pain medications, including Percocet or Vicodin, and found in many over-the-counter preparations. You should not exceed 4000 mg of acetaminophen in one 24 hour period.
- The patient should not lift anything greater than 10 pounds for 6 weeks.
- Contact Information If you experience a medical emergency, please call your local emergency response telephone number. For other questions, call the Oakcrest Medical Center 24-hour telephone number: 256-705-7296. Ask to be connected to the following service: Cedar Manor and Back Center.
- You have a surgical incision, which should be monitored for signs and symptoms of infection. These include, but are not limited to fever, increased wound pain, redness, swelling or discharge from the wound site. Should you note any of these, you should seek immediate medical attention, and contact Kerry Harper service via the MCHC operator.

- You may shower, but should avoid prolonged exposure of your wound to water for 6 weeks. Do not apply any ointments, salves or dressings to the wound, which should be left open to air.
- Your wound closures (staples/sutures) may be removed no sooner than 14 days following your operation. This can be accomplished by your primary care provider.
- You are discharged on narcotic medication. Do not drive or operate machinery while using this medication. Narcotics may cause constipation. You should increase your dietary fiber while using this medication. Over-the-counter stool softeners may also be helpful for constipation.
- Narcotic pain medication may cause constipation. To prevent this it is recommended that you drink lots of fluid and increase the fiber in your diet. Walking is a good way to get your bowels moving. If needed, you may take an over-the-counter stool softener such as docusate sodium (Colace) or senna (Senokot).
- The patient may be evaluated by an outpatient physical therapist and may use this dismissal summary as a prescription for the therapist to evaluate and treat the patient accordingly.

Please *do not* use DUPLICATE for text not listed above. If there's a duplicate section and it's not listed, contact your supervisor to get it added to the list.

13 Glossary

Default Value

All EVENT, TLINK, ALINK and TIMEX3 annotations have additional characteristics. If no other option is selected for these characteristics, the default values will be assumed and added.

Document Time/DOCTIME

The date of the note being annotated, usually at the top of the note. This is marked with DOCTIME.

Knowtator

Knowtator is the annotation tool to be used in this project

Span

An amount of text which is associated with an event. Spans do not have to be continuous, and in the sentence "Mr. X has experienced debilitating seizures", the span would be "experienced ... seizures"

Temporal Relations

Temporal relations is the relation between different events, actions, states and fixed timepoints in terms of time. If something happened before, after, during or around something else (or a fixed point in time), then those two things are temporally related, and that relationship must, in this schema, be annotated.

TimeML

This annotation schema is based on a system for marking the temporal relations of different words in a text called TimeML, developed by the TimeML working group. Please see <http://timeml.org/site/index.html> for more information.

14 References

References

- [1] James Pustejovsky, Kiyong Lee, Harry Bunt, and Laurent Romary, *Iso-timeml: An international standard for semantic annotation*, Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10) (Valletta, Malta.), European Language Resources Association (ELRA), 2010.
- [2] James Pustejovsky and Amber Stubbs, *Increasing Informativeness in Temporal Annotation*, Linguistic Annotation Workshop, 2011, pp. 152–160.
- [3] GK Savova, S Bethard, Will Styler, J Martin, and M Palmer, *Towards temporal relation discovery from the clinical narrative*, (2009), 445 (English).

15 Document Revision History (A Brief History of THYME)

- 1/29/14 - Added 12.1, discussing duplicate text.
- 7/31/13 - Clarified the interactions with NEGation and CONTAINS relations in Section 11.2
- 7/22/13 - Clarified CONTAINS, QUANTIFIER, added examples for BEGINS-ON, ENDS-ON, tweaked the "When to TLINK" rules to clarify that CONTAINS should only be used with containment.
- 6/7/13 - Added examples, added one reference to "point in time *and space*" in DocTimeRel to harden the guideline against inconsistencies due to special relativity.
- 5/11/13 - Added a small section (3.2.4) discussing specifically prohibited EVENTS (numbers and anatomical sites), added a small discussion of double-negation for NEG, discussed hedging in further depth.
- 4/22/13 - Added a great PREPOSTEXP example.
- 4/13/13 - Revised TIMEX3 span guidelines to match ISO TimeML standards, revised all example sentences.
- 3/26/13 - Moved "Recent Changes" section to its own section at the end of document to avoid clutter.
- 3/26/13 - Added example contrasting polarity negation with "negative" test results to 3.5.2
- 3/25/13 - Added 11.1 'Avoiding Inference', Fixed Typos, added another tricky TLINK example
- 3/19/13 to 3/23/13 - **Major Update:** Whole document re-examined to ensure conformity with finalized schema, improve pedagogical value for training.
 - Updated entire document to reflect head-word spans for EVENTS
 - Typographic Tweaks
 - Clarified AJCC Staging Code Spans
 - Added 5.1, "Why Narrative Containers", in preparation for a LAW paper and to clarify for release.
 - Added 4.1, explicit discussion of TIMEX3 spans.
 - Added 6.3, discussion of our point-algebraic understanding of our TLINKs.
 - Added paragraph on Latin names to 3.2.3
 - Added 6.1.6, discussion of TLINKing two TIMEX3s
 - Added Pedagogically-motivated "Caution" boxes to reinforce some points
 - Improved the "Schema Description" section to be a single-page resource for annotators to print.
 - Added examples to 5.4 showing the need for explicit TLINKing among EVENT narrative container anchors.
- 3/2/13 - Added "confirmed" examples to EVIDENTIAL
- 2/26/13 - Added discussion of EVIDENTIAL type
- 1/21/13 - Added discussion of latin names
- 1/15/13 - Added discussion of test results (under 3.2.3)x2x, erroneous small-caps
- 11/26/12 - Removed Protege Guidelines (Hallelujah!) to their own document, made a few small tweaks.
- 9/17/12 - Changed many examples to conform to new noun span policy
- 9/2/12 - Added discussion of implicit EVENTS
- 8/31/12 - Added a wonderful example of "pre-prandial"
- 8/15/12 - Removed AFTER as a TLINK type, clarified definitions of BEFORE, BEGINS-ON.

- 8/6/12 - Clarified TLINK section slightly, removed a verbal trainwreck in the pathology note section.
- 7/26/12 - Added section discussing Pathology note annotation
- 7/19/12 - **Major Update:** Reformatted all examples as numbered examples. This involved the one-by-one checking and revision of every example in the text, the addition of several more, and the clarification of further. Added discussion of the new, better way of representing links where needed. All examples and sections now reflect accurately the state of the entirety of the guidelines.
- 7/18/12 - Split off Nominal vs. Adjectival EVENTS, Updated adjective EVENT guidelines to reflect our use of only the head adjective.
- Early July 2012 - Added explicit discussion of marking EVENTS not required for treatment, but required for legal establishment of standards and practice (3.1.1). Added tricky TLINK discussion section. Added guidelines for choosing the best narrative container anchor, discussed the possibility of single-bounded narrative containers (with BEGINS_ON, etc). Removed all references to Knowtator in the main text. That felt good.
- 7/5/12 - Added the GENERIC modality, added 10.1 discussing discussions and generalizations, rules for TLINKing in those cases.
- 6/28/12 - Added discussion of nested narrative containers
- 6/20/12 - Discussed examples of negation with verbal and nominal EVENTS, dealing with "Medical History" sections.
- 6/8/12 - Addressed many of the "known issues". Added discussion of the (lack of) interaction between polarity and contextual modality. Discussed the nature of the split task (entities/relations), and discussed what to do in the presence of EVENT preannotations. Discussed the usage of the DATE type of TIMEX3 for more generic temporal expressions like "in the past" or "lately". Added a section to Special Cases and Constructions dealing with "prior" and "at the same time".
- 6/7/12 - Added discussion of family history when no explicit usage of the term occurs, revised that section to better discuss narrative containers in that context.
- 5/30/12 - Added example of "hold" being used as an aspectual EVENT, related ALINK.
- 5/14/12 - Added pre-introduction pagebreak, related quotation, discussion of Start Date vs. Rev Date
- 3/14/12 - Explicitly stated that all TLINKs between tests and their results must be of type CONTAINS. Added an example to support this. Also discussed handling (in 6.1.2) of tests, their results, and their date of occurrence. Finally added a few more ALINK examples.
- 2/22/12 - Discussed the aspectuality (or not) of the word "recurrence".
- 2/16/12 - Explicitly banned cross-sentence ALINKs.
- 2/11/12 - Removed four now-unneeded references to Transition/State/Process from the guidelines. Added an example to HEDGED based on an annotator question. Specified that TIMEX3s of type SET should always use TLINK type OVERLAP.
- 1/30/12 - Clarified the usage of HYPOTHETICAL when discussing future EVENTS, cleaned up OVERLAP and CONTAINS examples to remove ambiguities
- 1/25/12 - Removed all EVENT types except "aspectual" (removing the difficult Transition/State/Process distinction), rewrote that portion of the guidelines, and added "N/A" as a type.
- 1/7/12 - Changed the guidelines title to reflect our new acronym, THYME
- 1/5/12 - Added a section discussing sections not to be annotated.
- 1/5/12 - Added "Narrative Containers" section, completely re-did the When-to-TLINK guidelines in hopes to better constrain the possible TLINK annotations.

- 12/07/11 - Added examples for History of, added "plan" to special cases. Also clarified the "What is not an EVENT" section, removing some typos and adding "unusual" and "necessary" discussion.
- 9/6/11 - Added SET to TIMEX3 types, discussed handling of things like "twice daily", revised a typo in DURATION
- 8/17/11 - Fixed broken Figure numbers (Thanks Dave!)
- 8/08/11 - Discussed the Pre- and Post- expression types for TIMEX3
- 8/03/11 - Added an outline of the schema as needs to be implemented for annotation
- 6/28/11 - Revised TLINK guidelines as well as clarifying that EVENTS without DocTimeRel are not EVENTS at all.
- 6/15/11 - Revised the handling of Pre- and Post- expressions, clarified handling of missing body parts.

NOT FOR REDISTRIBUTION

16 Appendix 1: Clinical Note Section Labels

- 20100 - Revision History
- 20101 - Referral Source
- 20102 - Chief Complaint/Reason for Visit
- 20103 - History of Present Illness
- 20104 - Current Medications
- 20105 - Allergies
- 20106 - System Review
- 20107 - Past Medical/Surgical History
- 20108 - Social History
- 20109 - Family History
- 20110 - Vital Signs
- 20111 - Physical Examination
- 20112 - Impression and Report and Plan
- 20113 - Diagnosis
- 20114 - Administrative
- 20115 - Special Instructions
- 20116 - Advance Directives
- 20117 - Service Actors
- 20118 - Immunizations
- 20119 - Admission Findings and Test Results
- 20120 - Problem Oriented Hosp. Course
- 20121 - Final Physical Examination
- 20122 - Adverse Reactions
- 20123 - Diet / Nutrition
- 20124 - Discharge Condition
- 20125 - Condition at Discharge
- 20126 - Ongoing Care Orders
- 20127 - Admission Physical Exam
- 20128 - Ongoing Care
- 20129 - Follow Up Agreements
- 20130 - PHF and CVI Dates
- 20133 - Admission Medications
- 20135 - Anticipated Problems and Interventions
- 20136 - Post-op Services

- 20137 - EMTALA Statement
- 20138 - Patient Education
- 20147 - Dismissal Medications

NOT FOR REDISTRIBUTION