

## Annotation Guidelines for BioMedICUS Sentence Modeling

### Revision History

Version	Authors	Description
20170420	EAL	Initial guidelines
20171023	EAL	Update to guidelines after schema evaluation
20171207	EAL	Addition of “Unsure” entity
20180402	EAL	Addition of guidelines for annotating sentences that span multiple lines
20180409	EAL	Addition of “Labels” guidelines

This document includes the schema and guidelines for annotating **sentence** information in clinical text with the brat rapid annotation tool (brat). This information will be used to train the sentence model for BioMedICUS.

### Table of Contents:

1. Overview of Annotation Schema
2. General Guidelines
3. Annotation Guidelines and Examples

### 1. Overview of Annotation Schema

Entity (1)
<ul style="list-style-type: none"><li>• Sentence<ul style="list-style-type: none"><li>○ Used to identify sentence boundaries in clinical text</li></ul></li><li>• Unsure<ul style="list-style-type: none"><li>○ Used to identify text where sentence boundaries are unclear in clinical text</li></ul></li></ul>

### 2. General Guidelines

- Annotate as much information in the sentence as possible, and include punctuation (“.”, “,”, etc. at end of a phrase)
- Use “Notes” field at bottom of “Add Comment” window for any notes or comments.
- Sentences that span multiple lines should be annotated using the “add fragment” functionality.

### 3. Annotation Guidelines and Examples – Sentences

#	Rule	Description	Sentence
1	Sentence Units	All text within a note should be classified within a sentence. Traditionally structured sentences, as well as fragments, section headers, lists, and tables should be considered sentences. When possible, capture punctuation that completes a sentence. A period should signal the end of a sentence, unless it is within a numerical value or abbreviation.	This is a 59-year-old gentleman who presents with a prior history of malignant neoplasm of the bladder.  He is allergic to penicillin.  Patient will need pulmonary rehab referral upon discharge from TCU.
2	Section Headers	Parent and child section headers should be considered sentences, when it is obvious they are marking the beginning of a note section. When it is unclear, mark as “Unsure.”	Assessment and Plan: Laboratory Data: Subjective: Objective:
3	Labels	Labels within sentences should be included in the entire sentence annotation.  Labels that cause a break in the sentence (label: sentence) should be annotated as section headers.	Run test Q: 4 hr. <i>--Annotate as one sentence. “Q:” does not denote a stop in the sentence.</i>  Back pain: She reports this is improved today. <i>--Annotate “Back pain:” as a sentence and the following text as a separate sentence.</i>
4	Vitals	Annotate the Vitals line as “Unsure.”	Pulse: 75 Temp: 99.7 Blood pressure: 128/60 resp. rate: 18 O2 sat: 99%
5	Exam Section	The Exam section should follow rules 1 and 7. If new lines separate sections, these should be considered new sentences.	Neck: Supple. Lungs: Clear to auscultation. Heart: Regular rate; s1, s2 normal. Abdomen: soft, nt, nd, with good bowel sounds. Extremities: No edema.
6	Fragments	Sentence fragments that lack punctuation but contain phrases or thoughts should be considered sentences.	Altered mental status -continue fluid resuscitation discussed with patient’s daughter – no escalation of cares
7	Abbreviations/Acronyms	Abbreviations and acronyms that contain periods should not signal the end of a sentence.	t.i.d. p.o. P.T.A.
8	New Lines	New lines and carriage returns can signal the end of a sentence if they constitute discrete sentences that omit punctuation. They can also be considered sentences if they are fragments where the idea is not continued in the next line. However, if the sentence spans more than one line, this should be marked as a single sentence.	Mechanism: s/p fall Known Injuries: Subacute on chronic subdural hematoma Right hip fracture
9	Lists	When a list contains a carriage return, each carriage return signifies a new sentence. When no carriage return is included and values are separated by commas, colons, or semicolons, the list should be considered as a single sentence.	Outpatient Medications: 1. Aspirin 81 mg p.o. daily. 2. Ativan 1 tab t.i.d. p.r.n. 3. Finasteride 5 mg p.o. daily. <i>--Each line would be considered a sentence.</i>
10	Punctuation that occurs within sentences	Colons Semicolons Commas	Orthopedics consult: WBAT to RLE, DVT prophylaxis. 4/1/17 – agitation and delirium.

		Em dashes Backslashes Periods within numerical values and abbreviations	P.T.A., living at home with wife.
11	Tables	All associated within a table, as well as the headers and legend, should be considered as Unsure annotations. Annotate the entire table as Unsure.	<ul style="list-style-type: none"> <li>asthma</li> <li>mother</li> </ul>
12	Note Headers/Footers	Annotate the entire header or footer when no actual English phrasing/fragments are included as Unsure. English language phrases and fragments should be annotated as sentences.	D: 04/19/2016 T: 04/19/2016 23:50 Name: John Smith MRN: 0000000 DOB: 01/01/2000 Account: 1111111 Admitted: 2222222 <ul style="list-style-type: none"> <li>Document: A23456789</li> </ul>
13	Formatting Elements	When formatting elements are included to break up sections, annotate these as Unsure.	Subjective: the patient presents with SOB. <hr/> Objective: Temp 98.6 BP 160/90  <i>--Annotate the dashes in the above example as a single sentence.</i>
14	Lab Results	Test type values should be annotated along with test name value. Any parsable English language should also be considered a sentence.	WBC 28.1 29.1 28.7 32.5 CBC 110.5 97.3 99.6 105.2  <i>--All WBC values are associated with WBC as a sentence. All CBC values are associated with CBC as a sentence.</i>
15	Unsure	Annotate any text for which you are unsure of sentence boundaries as "Unsure."	