2011 i2b2/VA Co-reference Annotation Guidelines for the Clinical Domain

Organizing Team Members:
Ozlem Uzuner
Tyler Forbush
Shuying Shen
Guergana Savova
Wendy Chapman
Cheryl Clark
Lynette Hirschman
John Aberdeen
Martha Palmer
Arrick Lanfranchi
Jiaping Zheng
Brett South

Adapted from ODIE guidelines authored by Guergana Savova and Wendy Chapman.

Version Tracking

Date	Changes	Author
December 24, 2010	Original Adaptation from ODIE Guidelines.	Ozlem Uzuner
December 29, 2010	Modifications to include more information specific for annotators.	Brett South
January 12, 2011	Edits	Ozlem Uzuner
January 14, 2011	Additional comments and examples from the MITRE team	Cheryl Clark, Lynette Hirschman, John Aberdeen
January 23, 2011	Edits and comments	Ozlem Uzuner
January 13-25, 2011	Additional examples including revisions to the original ODIE guidelines to add additional examples and explanatory language for annotators.	Tyler Forbush, Shuying Shen, Brett South
January 26, 2011	Editing temporality according to organizer phone discussion	Tyler Forbush,
Feb 10, 2011	Insertion of PHI Class	Tyler Forbush

Co-reference Annotation Guidelines for the Clinical Domain

1. General information and background

These guidelines describe the most relevant portions of the ODIE guidelines adapted for the 2011 i2b2/VA Challenge. ODIE guidelines were based on the MUC-7 Coreference Task Definition (MUC-7, 1997).

Anaphoric relations (a.k.a. anaphoricity) in this task are relations between concepts where the interpretation of one of the concepts (the anaphor) relies on the interpretation of another concept (the antecedent).

Co-referential relations (or co-reference) are anaphoric relations of type identity. Two expressions are co-referential if they refer to one and the same. The focus of the 2011 i2b2/VA challenge is co-referential relations between concepts. The concepts in this task could be entities related to a patient, e.g., medications, or states describing the patient, e.g., pain. These concepts are common in medical text and resolving their co-reference presents problems to current NLP systems in creating a full view of the clinical situation. For example, if a system extracted "it resolved his pain immediately", that phrase may be useless unless the system can also distinguish what "it" was that resolved the pain, and which pain "it" resolved. These types of relations are present in the 3 concept classes from Challenge 2010 (Problem, Test, Treatment), as well as a new "Person" class that will be added this year. Concepts will be linked as co-referential pairs, and those pairs may then be linked to form a chain of concepts that are co-referential. The pairs A-B, B-C, C-D thus create a continuous chain from A to D. For example, a patient's name and other pronouns may occur throughout a document, and through a series of related pairs we form a chain of concepts that all refer to the patient.

Data for the 2011 i2b2/VA challenge comes in the form of two corpora. One corpus, created under ODIE project, contains clinical notes and pathology reports from Mayo Clinic as well as discharge summaries, progress notes, ED reports, radiology reports, surgical pathology reports, and progress notes from University of Pittsburgh Medical Center. The ODIE corpus was annotated for co-reference under the ODIE guidelines, which are separate from these guidelines.

The second corpus for the 2011 challenge includes discharge summaries from Partners HealthCare and Beth Israel Deaconess Medical Center (MIMIC II Database), and University of Pittsburgh Medical Center. This second corpus was annotated according to the guidelines presented herein. All clinical documents have been fully de-identified.

Annotation tool:
For annotation tasks we will use Knowtator version 1.9 beta 2 available on sourceforge http://sourceforge.net/projects/knowtator/files/Knowtator. The Knowtator tool is written for Protege 3.3.1. This version of Knowtator includes specific modifications to make the annotation task faster and more efficient.

2. Scope of 2011 i2b2/VA Challenge

Our task is to annotate co-referential relations within a given document and across the paragraphs and the sections of that document. Cross-document co-reference is out-of-scope for this challenge task.

Concepts for coreference relations are those annotated in the 2010 i2b2 corpus, with the new addition of person names and all pronouns.

The annotation task for 2011 is to:

a. Annotate mentions of persons using the "Person" class, including proper names (ie. John Smith), salutations (ie. Dr. Smart), and other references to a person (ie. her daughter, your doctor, the cardiologist, his nurse, etc).
b. Mark co-reference relationships between annotated Problem, Treatment, Test, and Person concepts found within the clinical document.

General Notes on these Guidelines: Because of the pre-annotations and complexity of examples in this guideline, we employ a color coding and numbering system in this guideline that is unrelated to the Knowtator schema.

Blue concepts are pre-annotated concepts from 2010 (Problem, Treatment, or Test).
Red concepts are new markables that must be marked if not already marked.
Overlapping markables are purple where the markables overlap. These may be red over other red markables, or red over blue markables.
In the coref section, coref pairs will share the same highlight color of green, yellow, or Blue.
Italic concepts are examples of exclusions and should not be annotated.
In examples, only the markables that are relevant to the example will be marked to provide clarity.

These guidelines provide "questions to ask" to help resolve annotator uncertainty and improve annotation consistency. Annotators should ask these questions about new markables or relationship links to ensure that annotations are consistent with the guidelines.

3. Markables
Markables for this task are the concepts from 2010 challenge as well as references to persons including pronouns and person names. The project includes several tasks relating to markables and creating new concepts.

3A. Problems, Treatments, Test from Challenge 2010
The Problem, Treatment, and Test concepts from Challenge 2010 will be provided to annotators as ground truth pre-annotations and CANNOT be modified or deleted.

Examples:
- The patient underwent hip replacement 5 days ago.
- She was initially started on Bactrim, but this was stopped because she was afebrile.
- Her CXR was ordered, which showed unilateral L pneumothorax.

For clarity, please read the Challenge 2010 guideline to better understand these classes and the span rules enforced in that effort.

3B. "Person" class added in Challenge 2011

Annotate all mentions of persons or groups of people. This class includes proper

names, personal pronouns, possessive pronouns, job titles, and groups. Markable spans should include salutations and trailing credentials. Note that white space token boundaries are used for this class.

Examples:

- Jane Smith presented for diabetes. Mrs. Smith was a poor historian. She could not remember her home medications well.
- ***Dr. [Name XXX]*** reviewed the case. He recommends the patient should remain on the ventilator until his condition stabilizes.
- Signed by: Joe Doe, MD
- Her pain has gotten worse, her doctor has increased her dosage.
- The nurse told the resident that the patient's condition has declined.
- The patient was transferred to *the Robert* for surgery by Dr. Butcher.
- This woman was gravely ill upon admission.
- The family was consulted, they support the treatment plan.
- Seen by cardiology, they recommended an echo in the morning.
- This is a 19 year old man presenting for broken leg.
- The patient has *Turner's* syndrome.
- His family was at his bedside.

Exclusions:

- Exclude descriptions or general mention of persons.
  - Jane Smith presented for diabetes. Mrs. Smith was *a poor historian*. She could not remember her home medications well.
  - Patient is *a 42 year old AA male*.
  - He gets aggravated by *people* easily.
  - This is *a 19 year old man* presenting for broken leg.

- Exclude proper names that were introduced during the De-identification process ONLY if they replace / refer to non-persons (see example below).
  - The patient was transferred to *the Robert* for surgery by Dr. Butcher.
  - The cancer of the left *Amanda* spread to lymph.

- Exclude proper names that are mentioned as eponyms of problems, tests, or treatments. These names will be included in the pre-annotated concept of other classes.
  - *Marfan* syndrome
  - *Parkinson's* disease

Personal pronouns will be provided as machine generated pre-annotations, and therefore may require you to add or modify. You may delete the pre-annotations only if they are not actually personal pronouns. Unlike ground truth concepts from 2010, you may modify or add new markables to the Person class in accordance with the guideline.

Question to ask: Is this token representative of a specific person or group of persons, and does the span include salutations and related information?

3C. "Pronoun" class, added in 2011
This class includes all pronouns that are not included in the person class, and they could refer to Problem, Test, Treatment or Person. Annotators can add pronouns, modify pre-annotated pronouns, and delete those that are not correctly identified as pronouns. Personal pronouns and those that refer to persons should be annotated using the Person class.
The most common pronouns will be provided as pre-annotations. In these situations the annotator may need to add, or modify instances in this class.

Examples of "Pronoun" Class:

- The pain, which was severe, required nerve block, which was a success.
- A CT was done, it showed bilateral ground glass appearance.
- The wound staples were left in. These can be removed by the patient.

Question to ask: Is this token a pronoun, and does this pronoun refer to a problem, treatment, or test concept markable?

3D. "Time Stamp" class
This class is used to annotate the first word of a document when the annotator begins work on the document, and the last word of a document when the annotator finishes that document.
Time stamps made using this class are required for each annotator on each document within the batch, and are not used to calculate batch time or billable hours. Time stamps are used to assess the elapsed time needed to complete this task on documents of different lengths and annotation density.
Because of the cognitive load and annotator memory needed to complete this task, annotators should start and finish each document without interruption. Any breaks or other interruptions should be made between documents.

3E. PHI Class
This class is used to annotate protected health information that the annotator feels has not been removed or resynthesized. We have pre-processed the documents using a tool to remove Protected Health Information (PHI). Some documents replace the PHI with a span similar to ***NAME (ZZZ)***. Other documents have realistic replacements with names and dates that seem realistic. Context and judgment should be used to decide what may be a replacement and what may be real PHI. Annotate any instances of missed PHI you may encounter as you review each clinical document.

The 18 HIPAA PHI categories are:
1. Names
2. All geographic subdivisions smaller than a state (street address, city, county, precinct)
3. For dates directly related to the individual, all elements of dates, except year. (date of birth, admission date, discharge date, date of death)
4. All ages over 89 or dates indicating such an age
5. Telephone number
6. Fax numbers
7. Email address
8. Social Security Number
9. Medical Record Number
10. Health Plan Number

11. Account numbers
12. Certificate or license numbers
13. Vehicle identification/serial numbers including license plate numbers
14. Device identification/serial numbers
15. Universal Resource Locators (URL's)
16. Internet Protocol addresses (IP's)
*17. Biometric identifiers
*18. Full face photographs and comparable images

4. Co-reference Pairs

The co-referent relationship slot links two concepts that are co-referential. Each concept can only be paired with at most one concept. Problem, Treatment, Test, and Person can only be co-referential to other concepts of their same class or Pronoun. Often concepts may be part of a set to subset relationship. These are not linked as pairs because they are outside of the scope of this guideline. Concepts should be paired with their nearest preceding co-referent concept.

These guidelines consider concepts as co-referential if they are of the same class, and refer to the same disease episode or event. Chronic diseases are considered the same disease episode regardless of temporality and are linked.

This task requires human annotators to use their knowledge and the document context to determine if concepts should be paired. In all cases annotators should use the "questions to ask" and the following rules to guide this decision:

- Paired concepts are limited to a single experiencer.
  - His father had lung cancer…. The patient was found to have non-small cell lung cancer.
    (The experiencers are different for "lung cancer" and "non-small cell lung cancer", so the concepts are not the same, thus are not paired)

- Clinical actions cannot be paired with devices or other concept types.
  - He underwent nephrostomy tube placement. The nephrostomy tube was removed the next day.
    (One is a clinical action, the other a device. These are not co-referent.)

- Synonyms and other variations, including the use of modifiers, are paired if they refer to the same event or episode..
  - The incision was clean and dry, we will continue to monitor the wound for signs of infection.
  - He had severe shortness of breath, but after resting for 10 minutes he had mild SOB.
    (In this context, the SOB has been one continuous episode from severe to mild and thus are co-referent. Separate epsisodes of SOB would not be linked in a pair)
  - The CXR revealed 8 mm obstructing stone… The renal stone was considered to be the cause of patient's symptoms… We recommended surgical procedure to remove ureteropelvic stone…
    (Though modifiers are different, these 3 markables are all referring to the

same kidney stone, and would be linked B-A, C-B.

- Do not pair concepts that take significant inferencing or assumptions for coreference.
  - The patient takes Tylenol at home for pain… The patient reports occasional abd pain.
    (Here we cannot be sure that the patient takes Tylenol for his abd pain without further information, they may not be the same pain.)

  - We suspect sepsis… The patient's UTI was treated with cipro.
    (The UTI may be urosepsis, but it is not clear that the UTI is the sepsis of which they speak)

- Pairing is limited to anatomic location if the location is explicated. Pairs must only link concepts or states of the same location. Context and clinical judgment should be used to link pairs in which location is explicated in one markable but not another.

  - The patient reports edema in his left leg. Exam revealed 2+ edema in patients hands.

  - The patient has a sternal scar. The CT shows pleural scarring, likely from past injury.

  - MVA resulted in 3 broken left ribs, and 1 broken right rib.

  - Right stone was removed surgically but the left stone was felt to need no intervention.

  - A right midclavicular incision was made… The incision was closed using Prolene.

    (Depending on context, these could be the same incision.)

- Brand name medications and their equivalent generic drugs are paired only if they are clearly of the same event..

  - The patient takes advil at home so we gave him ibuprofen 800 mg.
    (The advil at home was a different event than the ibuprofen given at the healthcare facility.)
  - Ancef was started for coverage, but we stopped the cefazolin after C&S came.
    (The Ancef and cefazolin are the same drug and of the same episode or action and are paired.)

- Pairs are limited by temporality or discrete clinical action.
  - The patient had a kidney stone in '00, he presents today with a left kidney stone.
  - Temp on admit was 40.1. Temperature today was 37.0
  - Hct 38.1. Her hematocrit came back and it was 41.0.

- Concepts with identical names may not always corefer. Context and judgment should be used.
  - Culture on blood sample showed NGTD. Culture obtained from foley bag, shows NGTD.
  - Family history: mother died of CAD 1999, sister also had CAD.

- Treatments are only paired if they are of the same episode and dosage.
  - 3 L NC on admission. Down to 1 L 02 today.
    (Because dosage has changed, these are considered two different treatments and are not paired)

- Diseases or diagnoses and their symptoms may not be coref pairs, depending on context.
  - Right hip pain is worse when it's cold. Was diagnosed with osteoarthritis.
    (The hip pain is a symptom of, but not identical to the arthritis)

- Treatment pairs are limited to the specified route of administration if it is explicated.
  - Morphine IR was replaced with IV morphine when patient was placed NPO.
  - Topical erythromycin applied to wound… Discharge medication: erythromycin 250 mg tab QID.

- Chronic diseases can be considered of the same episode and be paired unless otherwise indicated in the text.
  - Her COPD was diagnosed in 2000, today she says her COPD is still mild.
    (Because COPD is a chronic disease, it can be assumed this has been one continual episode and the COPD concepts are paired.)
  - Patient reports 10 year history of diabetes, his diabetes is well controlled today.

Because possessive pronouns are markables and they were included as concept modifiers in 2010, there will be overlapping concepts of Person class on top of other markables. These present a unique challenge in Knowtator when making pairs. To prevent missed co-referent pairs, annotators should always link the pronoun to its antecedent BEFORE resolving coreference of the overlapping concept. Overlapping concepts of Person on top of other person concepts are of particular concern and annotators must be vigilant in detecting and properly pairing each of the two concepts.

Person over other markable:
The rash on her hand is not painful, but her rash has spread to the entire arm.
(In this example, the "her" of her rash is a person markable that is layered on top of the "her rash" ground truth pre-annotation. The 2 "her" markables would be linked as a pair because they all refer to the same person. Always link the person pair first. "Her rash" would then be paired to "the rash".)

Person over person markable:
Her doctor recommended a nasal swab, which he then read and gave her a Z-pack.
("Her" is a person markable as well as "her doctor". Always ensure that both are paired

according to the guidelines. “Her” would be linked to “her” and “her doctor” would be linked to “he”.)

The person class presents unique challenges because of the De-ID system. Use context and judgment to determine if ***Name [XXX]*** ***Name [YYY]*** are coreferential to ***Name [XXX, YYY]***. In this case ***Name [XXX]*** would be paired with ***Name [YYY]***, and ***Name [YYY]*** would be paired with ***Name [XXX, YYY]*** to form a continuous chain of coref.

5. Chains

Annotators create coref pairs according to these guidelines, and post processing will transform the series of pairs into a coref chain. Coref pairs created by annotators will be programmatically marked and ordered to form a continuous chain of concepts. The concepts pairs of A-B, B-C, C-D would thus be ordered to form a continuous chain of concepts from A to D. Any out of order pairs such as A-B, C-A, D-B would be reordered programmatically to ensure the same continuous chain is created.

Be mindful of this output when creating coref pairs. In the above example, if A and D are not coreferential, an error has been made in the pairing process.

6. Slot Values

There are 4 slots on the right panel of knowtator.

1. Coref: When a co-reference pair is made, the other member of the pair is shown in this slot. This slot is limited to 0 or 1 members, thus you cannot coref the markable to more than one concept.
2. Uncertainty: When the annotator is uncertain about the pairing of the coref, they may make the pair and then select “uncertain”. It is important that the “uncertain” is selected on the correct markable. For example, A, B, C, and D may be co-referent. You are sure that B=A and C=B, but unsure if D=C. You would select the D markable, link it to C, and then mark the uncertainty on the D markable.
3. Normalized Information: This slot is for informational purposes only and is cannot be edited. This slot contains the UMLS description for the concept, which may be helpful in determining pairing.
4. Helpful: Annotators should select this slot if they read and found the normalized information useful in understanding the concept or determining pairing.

7. Annotation and Arbitration

Each batch of documents will be annotated by two independent human annotators. The merged document batches will then will then undergo arbitration by a third annotator. The opinion of clinician reviewers will be used to resolve particularly difficult disagreements.

Machine methods may also be used to correct common span errors and to create continuous chains from out of order pairs.

References:

Bagga, Amit. 1998. Evaluation of coreferences and coreference resolution system. In Proc. of the First Language Resource and Evaluation Conference. May 1998.

Bodenrieder, Olivier and McCray, Alexa. 2003. Exploring semantic groups through visual approaches. Journal of Biomedical Informatics 36 (2203) 414-432.

Borthen, Kaja. 2004. Predicative NPs and the annotation of reference chains. COLING-2004.

Borthen, Kaja. Last accessed June 20, 2008. Annotation scheme for BREDT. http://bredt.uib.no/publikationer/bredt_annotation.pdf

Castano J., Zhang J. and Pustejovsky J. 2002. Anaphora resolution in biomedical literature. International Symposium on Reference Resolution, 2002.

MUC-7. 1997. MUC-7 Coreference task definition. http://www-nlpir.nist.gov/related_projects/muc/proceedings/co_task.html

Olsson, Fredrik. 2004. A survey of machine learning for reference resolution in textual discourse. SICS Technical Report T2004:02. Swedish Institute of Computer Science ISSN 1100-3154.

Yang, Xiaofeng; Su, Jian; Lang, Jun; Tan, Chew Lim; Liu, Ting; Li, Sheng. 2008. An entity mention model for coreference resolution with inductive logic programming. Proc. of ACL-08: HLT, pages 843-851, Columbus, Ohio, USA, June 2008.