

CSCI-UA.0480-057

Homework Number 1

Due at Midnight after the 3rd Class

1. Download the annotation homework: [hw1.zip](#) Extract the files into a single directory (the program will not work if they are not in the same directory). Note that the Mae annotation tool was downloaded from github. Version 2.2.10 is included in the zip file (I downloaded it from the [mae github site](#)).
2. Follow the remaining instructions to create your annotated file and submit that file to GradeScope. Please follow these guidelines for submission:
 1. Name the file that you submit: "submission.xml".
 2. Submit it to the class Gradescope Page (posted on the main class website).
3. Run the MAE program to complete the annotation
 1. Load Mae
 1. From a UNIX command line, type “*java -jar mae-2.2.7-fatjar.jar*”
 2. Or (left) double click on the “mae-2.2.7-fatjar.jar” file
 2. Load the dtd file:
 1. File → New Task Definition → adjective.dtd
 2. Note that you can write your own dtd file to run MAE with a different annotation task
 3. Load the file to annotate:
 1. File → Open Document → state_of_the_union.txt
 2. If you load AM_state_of_the_union.txt instead, the first 2 paragraphs will have the instructor's annotation already there (as examples).
 3. Click on the **Adjective tab** so you see all the adjective features (otherwise, you will only be able to complete part of the annotation task)
 4. To mark an adjective:
 1. Drag left mouse over adjective and click: this will highlight the text you are selecting
 2. Right click on the highlighted text
 1. If you are an Apple user with single button mouse, this is equivalent to Ctrl+click

3. Move the mouse over “Create tag with extended text” and you will be given one choice: Adjective, which you should select with a left click
4. Change attributes in list of adjectives (left click and select on slot)
5. Save your work frequently
 1. File → Save Annotation as XML → submission.xml
 2. You can call your file what you wish, but the version that you submit to gradescope should be called 'submission.xml'
6. Exit when you are done (and and have saved your work)
4. Specifications for Annotation (More specifications can be found in the [Penn Treebank POS Guidelines](#), especially pages 13-17 -- the Penn tags for adjectives are JJ, JJR and JJS). Also see slides 40–42 from [Lecture1](#).
 1. Adjectives occur in two main positions
 - Attributive adjectives precede nouns that they modify
 1. Ex: *the **big** sandwich*
 - Predicative adjectives are linked to a noun phrase by predication
 1. Ex: *The sandwich is **big***
 2. Ex: *I made the sandwich **big***
 3. Ex: *Philosophy may seem **difficult** to understand*
 2. *Adjectives can have three morphological forms. Only mark adjectives as comparative or superlative if they are morphologically marked as such (i.e., if they have an -er or -est ending). For example, "more angry" has a comparative meaning, but the word "angry" does not have a comparative marking on it. So it should be tagged as **Normal**. Remember you are marking words and not phrases. Another processor which marks phrases would interpret the combination of "more" + "angry" as having a comparative meaning in combination.*
 - Normal: **big**
 - Comparative: **bigger**
 - Superlative: **biggest**
 - This is a **morphological** classification of a **word** which pertains to suffixes and prefixes.
 - It is not a **semantic** classification and it is not a classification of word sequences (more than 1 word)
 3. *Adjectives should not be confused with nouns, e.g., truck is not an adjective in the following example*
 - Ex: *The **truck** salesman*
 4. *Nouns, not adjectives can:*
 - occur in the plural, e.g., **trucks**
 - be modified by determiners like the or a, e.g., **a truck, the truck**

5. *Determiners are NOT adjectives (they occur before adjectives). The following list of words are determiners and they should not be marked as adjectives: **such several one most more many less few enough both all your those this these their the that some our no neither my its his her every either each any another an a***
6. *Cardinal Numbers ARE NOT adjectives: **one, two, three, ...** (they are determiners)*
7. *Ordinal Numbers ARE adjectives: **first, second, third, ...***
8. *Difficult cases:*
 - *A word can have distinct meanings as an adjective and as a noun*
 1. ***They are studying for the final.** (Noun)*
 2. ***This was their final attempt.** (Adjective)*
 - *Adjectives may be used as nouns with an adjective + one meaning*
 1. ***They exploit the poor.** (Don't mark these as adjectives)*
 2. *Here **poor** means something like **poor people***
 3. *Leaving problem cases like these out of the specifications would lead to inconsistency and noisy data. Thus there are 2 possible choices:*
 1. *The specifications should state that the these should always be marked as adjectives*
 2. *Or that they should never be marked these as adjectives.*
 4. ***Poor** is almost always an adjective. These noun-ish cases are very rare. This may be a reason to choose "always mark them as adjectives"*
 5. *Compare with predicative position, e.g., **They are poor.** This is a clear adjective case. If **poor** was a noun here, the plural form would be used (**They are poors**).*
 - *Frequency is an issue, e.g., assume color words are adjectives because their noun-like uses are rare*
 - ***I really love this red** (noun) vs **That clown nose is red** (adj) or **The red nose***
 - *We are assuming that present-participle and past-participle verbs should be marked as adjectives when they occur attributively (before the noun), e.g., "**rotating**" in the sentence "**Green rotating objects appeared on the screen**"; and "**covered**" in "**The big covered wagon traveled 40 miles in one day**". When these same words occur predicatively, we are assuming that they are verbs, e.g., "**The object was rotating.**" and "**The wagon was covered by the inspector.**"*
 - *There are some participles which can be ambiguous in predicative position and for these it may be more difficult to figure out whether they are an adjectives or verbs. For example, "**John was understanding the***

*material in the course with difficulty" vs. "John was very understanding". In these cases, you will have to figure out which sense is being used. The verbal sense of **understanding** implies a thinking process, but the adjective implies that the person is **kind**. There are many adjectives that look like participles such that can be distinguished by sense. Testing them with "very" which only occurs with adjectives or adverbs (**slowly, quickly**) or phrases that tend to occur with verbs (**with difficulty**) is a good way of disambiguating.*

- *Adjectives can be part of idiomatic or nearly-idiomatic constructions involving verb-noun combinations. If the word does not have any other obvious part of speech (noun, verb, etc.) and it maintains its adjective-meaning, assume that it is an adjective. In most such cases, assume that it is predicative.*

1. Mary fell ill

2. John went ballistic

- *Assume that idioms like "**red herring**" are marked with their non-idiomatic parts of speech, i.e., "**red**" would be an adjective. These idioms, unlike the previous examples, consist of standard sequences of parts of speech, but have idiosyncratic meanings. The previous examples have unexpected part of speech sequences.*

9. *Adjectives can be modified by other words: **light red, very hungry, quite upset***

10. *In attributive position, adjectives occur after the determiners and before any nouns, e.g.,*

- ***the hairy mountain gorilla***
the = determiner; hairy = adjective; mountain = noun; gorilla = noun

11. *Grading of Annotation Homework*

- *The final scoring from 1-10 will be determined on a competitive basis*
- *We will generate the following scores for analysis:*
 - *Precision, Recall and F-measure against the Professor's annotation (which is usually approximately the same as previous class's pooled results)*
 - *Precision, Recall and F-measure may also be measured against some Pooled annotation results. In other words, each word in the corpus will be marked as an adjective in the pooled results, iff 50% or more of the students marked it as an adjective .*
- *We will also discuss the "difficult" cases*
 - *Cases that the Professor's and the pooled results differ:*
 1. *An error in the Professor's annotation?*
 2. *An unclear part of the adjective specifications?*
 - *Other cases that students want to ask questions about*

Accessibility